
P is psychology. P is positive. P is passion. P is pace maker.
P
discover inside...
[美] 罗伯特·索尔索（Robert L.Solso）（美国内华达大学）M·金伯利·麦克林（M.Kimberly MacLin）（美国北艾奥瓦大学） 著
张学民 周文斌 郑亚芹 译
EXPERIMENTAL PSYCHOLOGY:A CASE APPROACH
索尔索 & 麦克林
（原书第8版）
实验心理学
全球采用率第一的实验心理学教材
被哈佛大学、麻省理工学院等200多所国内外知名大学采用
已故心理学大师，美国心理学协会创始人索尔索博士经典之作
国内实验心理学顶尖专家杨治良教授专文推荐
中国人民大学出版社
China Renmin University Press

著作权合同登记号
图字：01-2008-2892

国内外200多所大学采用，其中包括哈佛大学、哥伦比亚大学、麻省理工学院、美国西北大学、北京师范大学、华东师范大学等知名大学和学院。

这本教材反映了一套匠心别具、独树一帜的教学方法。本书在介绍基本原则时，穿插了大量案例，这使得原则的阐释生动、形象，不再枯燥或流于形式。第二部分的案例涉及面广，贴近生活，趣味盎然，深受师生喜爱。另外，本教材是第8次修订版，可谓是精雕细琢之上品。

——杨治良，华东师范大学心理系教授

这本教材的内容通俗易懂，呈现方式生动，使读者能更容易地学习和理解比较艰涩的实验心理学课程，能够在主动学习的基础上达到入门、理解和掌握实验心理学的基本原理、思路、实验设计方法和过程的目的。本书独到的写作风格和教学实践思路非常值得我们借鉴。

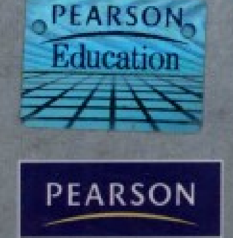
——张学民，北京师范大学心理学院副教授

我教授实验心理学已经12年了，一直都采用这本教材。这本书的内容中肯而全面，精彩之处在于第二部分的15篇实验报告。我认为让学生在案例基础上理解概念和原理是至关重要的，所以，我非常推崇这本教材。

——巴顿·鲍尔森，美国西北大学

欢迎老师登录百度贴吧“湛庐教师中心”吧讨论

欢迎登录湛庐教材教师服务中心www.cheersbooks.com，点击“湛庐教材”



总策划人：韩　焱
执行策划：郭咏雪
营销总监：张晓卿
设计总监：夏红卫
责任编辑：黄珏苹
封面设计：湛庐文化　梁庆博
投稿请至：service@cheersbooks.com



PEARSON

上架指导：心理学/实验心理学



ISBN 978-7-300-10641-0/B·636
定价：52.00元

图书在版编目（CIP）数据

索尔索 & 麦克林实验心理学（原书第 8 版）/（美）索尔索，（美）麦克林著；张学民等译．
北京：中国人民大学出版社，2009
ISBN 978-7-300-10641-0

Ⅰ. 索…
Ⅱ. ①索… ②麦… ③张…
Ⅲ. 实验心理学—研究
Ⅳ. B84

中国版本图书馆 CIP 数据核字（2009）第 067090 号

索尔索 & 麦克林实验心理学（原书第 8 版）

［美］罗伯特 · 索尔索　M · 金伯利 · 麦克林　著
张学民　周义斌　郑亚芹　译

出版发行	中国人民大学出版社		
社　址	北京中关村大街 31 号	**邮政编码**	100080
电　话	010－62511242（总编室）		010－62511398（质管部）
	010－82501766（邮购部）		010－62514148（门市部）
	010－62515195（发行公司）		010－62515275（盗版举报）
网　址	http:// www. crup. com. cn		
	http:// www. ttrnet. com（人大教研网）		
经　销	新华书店		
印　刷	北京京北印刷有限公司		
规　格	214mm × 275 mm　16 开本	**版　次**	2009 年 6 月第 1 版
印　张	15.5 插页 4	**印　次**	2009 年 6 月第 1 次印刷
字　数	428 000	**定　价**	52.00 元

索尔索&麦克林实验心理学

（原书第8版）

Experimental Psychology: a Case Approach

（美） 罗伯特·索尔索（Robert L. Solso）
M·金伯利·麦克林（M. Kimberly MacLin） 著

中国人民大学出版社

目录 CONTENTS

第一部分　实验心理学的基本原则

第1章　科学研究概述 2

1.1　实验心理学的主题 2
1.2　科学 3
1.3　事实、理论和推测 5
1.4　心理科学 6
1.5　科学方法 6
1.6　实验设计 7
1.7　科学方法的假设在实验心理学中的应用 7
1.8　心理学实验 10
1.9　实验心理学的假设提出 10
1.10　其他研究方法 11
1.11　观察法 11
1.12　档案研究 15
1.13　相关研究 16
1.14　调查研究 17
1.15　结束语 18

第2章　心理学文献：理解性阅读并获得研究思路 19

2.1　思路、灵感和心理学文献 19
2.2　使用因特网 20
2.3　电子数据库：索引、摘要和全文 20
2.4　心理学文献阅读与理解 22
2.5　文献阅读记录 22
2.6　结束语 22

第3章 心理学基本实验设计 24

3.1 实验心理学的逻辑 24
3.2 操作性定义 25
3.3 自变量和因变量 26
3.4 实验组和控制组 29
3.5 结束语 32

第4章 高级设计技术 33

4.1 因子设计 33
4.2 准实验设计 40
4.3 功能设计 42
4.4 结果的泛化 45
4.5 结束语 45

第5章 实验设计与控制 46

5.1 控制的比较 46
5.2 科学推论 46
5.3 控制类型 46
5.4 实验范式 48
5.5 处理组和控制组的使用 51
5.6 结束语 56

第6章 被试变量控制 57

6.1 处理组被试的同质 57
6.2 随机分派 58
6.3 被试匹配 59
6.4 被试内控制 61
6.5 被试缺失 62
6.6 结束语 63

第7章 实验设计的评论 64

7.1 实验概要........ 65
7.2 结束语........ 69

第8章 实验研究的伦理问题 70

8.1 心理学家的伦理道德和行为规范........ 70
8.2 使用人类被试的研究........ 73
8.3 案例分析........ 75
8.4 结束语........ 77

第9章 研究过程 78

9.1 如何研究........ 78
9.2 撰写研究论文........ 85
9.3 数据报告........ 89
9.4 样稿........ 92
9.5 投稿........ 98
9.6 专业会议的展板........ 100
9.7 结束语........ 101

第二部分 实验分析

第10章 品尝可乐 106

是不是喝可乐越多，就越能准确分辨不同品牌的可乐？（控制额外变量）

10.1 引言........ 106
10.2 分析........ 108
10.3 问题........ 110

第 11章　停车场的领地 112

狂按喇叭能让占着车位的司机尽快走人吗？（现场研究）

11.1　引言 112
11.2　问题 120

第 12 章　旧情难忘 121

是不是不去回想旧情人，就可以免受失恋的折磨？（皮电测量研究）

12.1　引言 121
12.2　分析 121
12.3　问题 124

第 13 章　图片记忆 125

理解有利于更好地识别和记忆无意义的图片吗？（自由回忆法与联想回忆法）

13.1　引言 125
13.2　分析 127
13.3　问题 132

第 14 章　荷尔蒙与玩具偏好 133

为什么男孩爱玩汽车，女孩爱玩洋娃娃？（被试变量的选择与控制）

14.1　引言 133
14.2　分析 136
14.3　问题 138

第 15 章　母性行为 139

伟大的母爱是如何产生的？（动物行为实验研究）

15.1　引言 139
15.2　分析 142
15.3　问题 144

第 16 章　儿童的推理 145
整个物体、物体的部分，哪个对儿童的认知更重要？（儿童心理实验研究）

16.1　引言 145
16.2　问题 150

第 17 章　聪明的海豚 151
如何把海豚变得越来越聪明？（小样本设计）

17.1　引言 151
17.2　分析 154
17.3　问题 160

第 18 章　转变角度 161
你记忆中的世界是真实的世界吗？（自我报告法或内省报告法）

18.1　引言 161
18.2　分析 162
18.3　问题 168

第 19 章　治疗愤怒 169
愤怒引发的抑郁可以治疗吗，效果好吗？（单被试研究）

19.1　引言 169
19.2　分析 170
19.3　问题 175

第 20 章　亲社会行为 176
电视节目能给孩子树立好的榜样吗？（场景性实验研究或模拟性实验研究）

20.1　引言 176
20.2　问题 180

嗨！您好，我是刀客网络营销系统创办人，方胜山。

请记着我的名字，因为在你未能通过互联网获得成功之前，我会一直陪伴着你！

访问下面链接获取更多刀客营销资讯:

刀客网络营销工作室简介：http://player.56.com/v_OTM5MzIwNDM.swf

刀客网络营销腾讯微博 http://t.qq.com/frederick_yan

刀客网络营销工作室官网 http://www.daoke123.com

您能理解“8 个小时之内求生存，8 个小时之外求发展”这句话的深刻意义吗？它告诉我们：

人的一生真正价值在于工作之外，工作只是为我们提供温饱，绝对不是生活的全部！

在这个物欲横流的现实社会环境下，人的绝大多数矛盾和需求都是由金钱的匮乏引起的，相信你一定在寻找一种能够快速赚钱的办法。因为足够的金钱，可以让你做到下面的几点：

1.丰富的物质生活；

2.家人生活的保障

第 21 章　酒精与避孕套 181

酒能助性还是乱性？（多重实验）

21.1　引言 181
21.2　分析 182
21.3　问题 191

第 22章　空手道技术 192

专家比新手能记住更多的动作吗？（操纵被试变量）

22.1　引言 192
22.2　分析 194
22.3　问题 195

第 23 章　在日本的纠纷调解 196

公平的结果和程序，哪个更重要？（跨文化研究）

23.1　引言 196
23.2　问题 199

第 24 章　错误的供罪 200

人们真会承认没有犯过的罪行吗？（实验中的欺瞒技术）

24.1　引言 200
24.2　分析 200
24.3　问题 203

附录 A　基础统计的计算程序 205

A.1　数据类型 205
A.2　使用附录A 206
A.3　集中趋势 207

A.4 差异量数 208
A.5 相关量数 208
A.6 方差分析 211
A.7 线性回归 212

附录B 统计表 **214**

术语表 **218**
译者后记 **225**

推荐序

从实际研究案例入门

杨治良
华东师范大学心理系教授、博士生导师

1879年，冯特在莱比锡大学建立了世界上第一个心理学实验室，标志着科学心理学的诞生。从那时起，实验心理学就奠定了其在心理学中的基石地位。一百多年来，心理学的范式和隐喻不断更迭，但主流心理学界始终将实验法作为探索大脑、心理和行为规律的最重要的方法。原因很简单，实验方法是探索事物间因果关系的最佳方法，而因果关系则反映了世间规律的最高层次。基于实验心理学这一不可替代的地位，它也是我国各高校心理学专业的核心课程之一。

美国实验心理学开展得比较早，积累也相对丰富，因此，引进一些相关教材，对我国实验心理学的科研和教学工作都是非常有益的。当前，我国出版界实际上已经翻译、影印了不少优秀的实验心理学教科书。但在我看来，这本由中国人民大学出版社组织引进，张学民教授领衔翻译的《索尔索 & 麦克林实验心理学》，并非仅仅是又一本国外教材这么简单。本次的引进工作大有裨益。由罗伯特·索尔索和M·金伯利·麦克林教授编撰的这本教材，实际上反映了一套匠心别具、独树一帜的教学方法。在我看来，这套教学方法主要有如下两个突出的特点。

首先，本书采用了所谓的"案例教学法"，通过对多达70多个研究案例的详细介绍，使读者能够在具体问题中理解抽象的实验心理学原理。本书第一部分在介绍实验心理学的基本原则时，就穿插大量案例，使原则的阐释生动、形象，不再枯燥或流于形式；而第二部分又专门选择了15篇较为完整的实验报告，并附以批判性的评论和问题，向学生展现出许多实际研究的真实面貌。作者这种"自下而上"的方法确实有利于学习者（特别是初学者）更快、更好地获得知识。

其次，本书选取的研究问题涉及面广、贴近生活，容易引起读者的兴趣。实验心理学如今已经渗透到了当代心理学的方方面面，不仅认知心理学家、动物心理学家对实验法感兴趣，社会心理学家、发展心理学家、临床心理学家、教育心理学家、法律心理学家等都在很大程度上依赖于实验方法。本书的案例就很好地反映了这一点。读者甚至会略感惊奇地看到，连"如何忘却旧日恋情"、"电视节目能否给孩子树立好的榜样"这样的问题也可以通过实验法来解答。相信通过阅读本书，读者会更加坚定用科学实验方法来解决重要的、有趣的实际生活问题的信心。

除了上述两个突出优点外，本书在内容上也十分完备：不仅囊括了最重要的研究原则，还介绍了文献检索与阅读、研究伦理、数据分析、投稿等研究者所面临的实际问题。因此，本书也可以起到实验研究参考手册的功用。

本书被国内外许多大学心理系广泛采用，深受师生喜爱，口碑甚佳。当前译本是该书的第8次修订版，可谓是精雕细琢之品。

实际上，我们华东师范大学心理系近年来也一直将该书作为实验心理学课程的重要参考书之一。

基于本书的突出优点、良好口碑以及我的亲身教学经验，我郑重将本书推荐给我国从事心理学教学和科研的广大学者和教师，以及心理学专业的研究生和本科生。相信通过阅读本书，大家能够扎实地掌握实验心理学知识和相关技能。当然，我更希望看到部分读者在未来做出优秀的心理学实验！

中文版序

亲爱的读者，我很高兴为《索尔索 & 麦克林实验心理学》写中文版序。

与这本书的结缘始于帮助我最喜爱的教授——索尔索教授编辑本书的第一版，那时，我还是他的学生。多年过去后，我超越了编辑的角色，成为了合著者。索尔索教授和我对此书倾注了大量心力。非常不幸的是，我的良师益友——索尔索教授于2005年去世了。为这本书写序，对我来说是悲喜交加。他的言语、他的精神以及他希望通过本书教导学生的热切渴望都体现在这本书中。

索尔索教授热爱中国，在他人生的最后几年，他曾多次到访中国。索尔索教授总是充满好奇，在职业生涯和个人生活中，他总在鼓励其他人要对周围世界充满好奇。这些年来，我们一直在努力解答使读者好奇的心理学问题。每一次修订这本书，我都会在心里想：读者需要什么？怎样能让读者更清楚地理解概念？什么样的案例更好？什么样的实验报告能引起读者的兴趣，并帮助他们理解实验设计的元素？我能把恼人的实验心理学写得有趣吗？我能把迷惑变为好奇吗？将通常意义的“有趣”和实验设计联系起来是颇具挑战性的。但事实上，心理学研究是非常有创意的，也是心理学引人入胜的一个方面。

对实验设计原则的学习可以帮你解答有关人类行为的各种问题，并为你理解和分析他人的研究提供工具，而且，你可以更理智地面对繁多的实验发现。我试图在本书中提供达到上述目的的所需信息，并希望你可以通过案例和实际的研究来学会这些原则。

是你们帮助这本书不断改进。读者说需要术语表，所以，我在本书中增加了术语表；读者提出某个案例或论文很枯燥、很晦涩，我就换掉。如果你喜欢这本书的某些方面，更重要的是，如果哪些方面你不喜欢，请与我联系，我非常想知道你们的反馈。我的邮件地址是：kim.maclin@uni.edu。

享受你的研究之旅吧！你将以全新的视角看待心理学。

Kim

作者简介



罗伯特 · 索尔索，原美国内华达大学里诺校区心理学系主任。于2005年1月16日逝世。1981年索尔索博士被选为富布莱特学者，他在各国讲学，为国际心理学研究做出了杰出贡献。

索尔索博士是美国心理学协会（APA）的创始人之一，也是西方心理学协会（WPA）的前任主席。他在内布拉斯加大学获得硕士学位，在圣路易斯大学获得博士学位，并在斯坦福大学从事博士后研究。他因心理学方面的突出贡献而获得了美国心理学协会的会长奖。

索尔索博士写过37本著作，80多篇文章，他编写的心理学教材被各国大学广泛采用，其中包括:《实验心理学》、《认知心理学》、《认知与视觉艺术》、《艺术心理及大脑意识的进化》以及《心智科学》等。



M · 金伯利 · 麦克林，美国北艾奥瓦大学心理学及法学副教授。她于内华达大学获得社会心理学博士学位。麦克林博士主要从事心理学概论、社会心理学、研究方法及心理学与法律课程的教学。她的研究活动在社会心理学、认知心理学及法律心理学间架起了桥梁。她注重将实验室研究成果应用于实际生活。

致教师

非常高兴《索尔索 & 麦克林实验心理学》这本书第八版出版了，在 20 世纪，实验心理学家涉足心理学的很多研究领域，当心理学已经成为了公认的正统学科，我们的教科书和研究论文也必然随之更新和改变。该版本的实验心理学包含了心理学的各个研究领域，并清楚详细地阐明了如何做心理学实验。

最初写这本书的目的是为了阐述实验心理学家们所运用的实验设计的基本原则。过去，实验心理学讲授和讨论的是繁琐的理论统计方法，或者将内容仅集中在某些特定的研究领域。但是，通过多年给学生上实验心理学的经验，以及对如何上好这门课程的探讨，我们总结出给学生讲授实例的方式，通过实例使学生掌握实验设计的基本原则。用认知心理学的术语讲，就是“自下而上”的方法，强调通过案例来掌握实验的原则。尽管这种方法是从具体到抽象，但我们的目的是对实验心理学的这两个方面都给予应有的重视。因为实例和原则对于实验心理学来说，是同样重要的。

本书的教学方法主要是采用具体的实验来帮助学生掌握如何将设计的原则应用于研究。在这个版本的实验心理学中，学生可以看到、评价和分析大约 75 个研究案例和实验，这些案例涉及实验心理学的各种设计原则以及可能遇到的问题。除了理解这些实验设计，学生还可以轻松地学习文献，并从心理学的这些资料中学到更多的东西。

无论是在高能物理、土木工程、会计、计算机程序设计，还是心理治疗、创作或细胞生物学等领域，案例教学一直是一种传统的教学方法，并被普遍采用。但是在心理学领域，在实验设计的课程中，这种实践因为一系列的哲学和理论问题而进展缓慢。在一些教育培养中被认为重要的东西，在相关的实际研究中，却变得很难应用（尤其是你最初开始学习这些内容）。我们认为案例教学是发展批判思维和研究实践的核心所在，尤其是对刚刚学习心理学的学生或是刚开始独立从事心理学研究的人更是如此。

我们是从指南的角度出发来写这本教材的，就像我们单独指导一个学生如何阅读文献。首先，我们会指出一个实验设计的基本原则和问题。然后展示心理学的研究文献是如何处理这些原则和问题的。我们还引用了已出版的文章，就像一个主讲教师坐下来和学生一起阅读一篇文献。通过从学生和指导教师的反馈情况来看，这种方法非常成功。

本书的第一部分概括描述了实验设计的基本原则。内容涉及当前的一些研究实例，并对这些很多教师在教学中采用的内容进行了更好地组织。尽管实验心理学是以实验室研究为主，但我们还是将研究方法的理论扩展到了传统实验心理学之外的学科领域。如组织研究者进行决策，论证他们在研究过程中的每个步骤，能够确定采用最恰当的方法来研究特定问题等也是学生需要掌握的。所以，本书还包括了另外一部分内容，那就是其他的研究方法。研究过程的相关章节得到了扩充和更新，包括最新的电子数据库信息、计划和实施研究、为研究找基金、写摘要、参加研讨会、准备研究成果展示、撰写专业论文等。

本书的第二部分内容包括 15 篇文献。这些文献是从心理学的主要领域中认真挑选

出来的，涵盖了工业心理学、认知心理学、社会心理学、动物和动物行为学研究、现实中的问题、跨文化研究、心理治疗、单一被试设计、教育心理学、行为矫正和儿童心理学以及其他领域。我们通过这些文章来阐述第一部分中的实验设计。我们还进一步阐明了在一些研究论文中，实验设计方面的特殊问题。这些问题包括实验控制的问题、特定领域的实验、被试选择、小样本实验、动物实验、临床研究，在实验室进行的社会行为研究以及其他方面的问题。我们发现，这种方式可以使学生学习和掌握心理学不同领域的实验设计，还可以学习同一领域不同形式的实验设计。一些教授告诉我们，无论是基本内容还是附加内容，第二部分的内容是他们为什么选择这本书作为教材的主要原因。

附录A和B包含了基本统计原理和统计表，这使本书可以被普遍应用于需要统计分析方法的课程中。附录还可以使教师论证本书大量统计检验的计算过程。它也是教师的统计检验手册，其中包括了统计检验的问题、课堂演示、讨论问题和讲解的方法。

致学生

我们为什么会在这里？你们在这门课上将做什么？通过这本书、这门课，你将学到我们是如何了解心理学的？你将学会如何做研究，如何加强你的阅读、思考和写作的技能。最起码可以掌握如何找到你心中的关于人类思想和行为问题的答案。

本书主要是阐述实验心理学的研究方法。其中相当部分内容是关于如何控制心理学实验和通过观察收集可靠的数据。我们用“案例入门”这个词是希望说明通过心理学专业文献中的实例来阐述心理学实验设计的不同原则。仔细学习这些实例，它们都是心理学主要研究领域中有代表性的设计严谨的实验，包括动物研究、儿童心理学、社会心理学、认知心理学和应用心理学以及其他领域。在本书中也特别强调了实验研究中的伦理道德问题，并且给出了如何提出好的研究设想、如何写研究报告、如何在专业会议上报告你的研究的一些基本原则。

第一部分

实验心理学的基本原则

科学，即解惑。——托马斯·库恩（Thomas S. Kuhn）

本书分为两个独立的部分。第一部分介绍实验心理学的基本原则，第二部分则对从心理学文献中选取的真实实验进行了分析。

本书首先进行心理学科学性质与方法论的探究，并通过研究实例对每一要点进行说明。第一部分内容包括实验设计的基本原则和开展研究需要注意的道德伦理问题，其中用大量篇幅分析了实验控制的问题，研究者正是通过实验控制的方法保证心理学实验的完整性。

第7章提供了一些实验研究的简要描述，其中每个研究概述都存在至少一个概念性或技术性的缺陷。当你阅读这些评论时，要努力找出其中的错误。这种练习有助于提高设计良好实验的能力。

第一部分还提供了关于伦理问题、研究思路和研究具体过程的知识。

第二部分呈现一些真实的心理学实验。每个实验案例都描述了一个或多个具体的设计或实验问题，有些案例中包含了详尽的分析（有几个案例留给学生评论）。学生可以参照第二部分的研究实例进行第一部分内容的学习，这有助于加深知识的理解。在完成两部分的学习之后，学生应具备了阅读、理解、分析、计划以及开展心理学实验研究的能力。

第1章　科学研究概述

一个人可以由于各种各样的原因而被科学吸引：有实现生存价值的渴望，有探索新领域的激情，有寻求秩序的希望，以及检验现有知识的动力。

——托马斯·库恩

心理学是不是科学？心理学与科学是毫不相干的词汇吗？心理学在“真正”科学的阴影下度过了太长的时间。直到今天，还有很多人认为心理学仍处于前科学阶段，心理学只是处于炼金术阶段，是对人类行为的复杂本质和大脑与思维的内在机制进行研究。那么，心理学可以真正被称为“科学”吗？当然可以！因为，科学的特征是其研究的方法，而非研究的内容。科学在于你的作为。与化学或生物学类似，心理学有传统的实验室，有实验的技术设备，有对数据的统计分析，以及由此得出的可靠结论。心理学的主题变化多端，但系统的实验探究是心理学与哲学、外行意见或纸上谈兵的区别所在。

如今，心理学家对思维和行为的实验研究遵循着其他科学领域所遵循的科学研究的基本原则。虽然心理学家的研究对象（思维和行为）与其他研究领域有所区别，但探究的方法是类似的。

1.1　实验心理学的主题

实验心理学家研究广泛的现象，例如情绪的心理成分，人们在快餐店的进餐行为、诗歌学习、不同身份个体的人际关系、苍枭的捕食行为、新生儿的眼动、人格特征、精神分裂症的病因、解决逻辑问题的大脑机制、面孔吸引力等。这些都是心理学家可以使用实验方法进行研究的合法问题。

外行人一般通过推测、主观评论或个人经验来思考上述主题。这种非科学的解释虽然有时看起来有趣，但通常是不可靠的，并可能得到错误的结论。例如，你希望了解接近自己感兴趣的对象的最佳搭讪方式，这能否通过实验进行研究呢？事实上，答案是：可以。克蓝克（Kleinke）、米克（Meeker）和坎宁安（Cunningham）等人都开展了关于如何开始交谈的研究。在克蓝克等人的一次调查中，被试对几种开场白进行了等级评定：“冷不冷？一起暖和暖和！”“你是学生吗？”“我有些害羞，但我希望能认识你。”这些开场白分别表示不同的社会接近类型：可爱－轻浮型、无害型和直接型。总体上，评定者认为可爱－轻浮型的开场白最不可取。

在这些常模数据的基础上，坎宁安进行了一次现场测试，他让主试在一个酒吧里使用这几种社会接近类型的开场白与多位异性搭讪。一般来说，女性对可爱－轻浮型的开场白反应消极，对直接或无害的开场白反应积极。而男性的反应比较模糊。或许当女性主动搭话时，无论是无害抑或可爱的方式，都会被男性知觉为积极的。该研究说明，即使是搭讪者也可以通过实验来进行分析。

是的，我们能够科学地对各种各样的人类思想与行为进行研究，即按照指定的方法进行研究并获得可靠且有效的结论。**信度**（reliability）和**效度**（validity）在心理科学中有特殊的意义。信

度是指实验、测量或观察的一致性。当研究者在某个情境下观测到某种行为，为了保证这一观测的可靠性，该研究者或其他研究者应当能够重复这次观测。效度是指某次观测或测量的真实准确性。例如，你开发了一种智力测验，并声称它能够预测学业成功与否，那么这个测验能够准确预测这种成功的程度就称做效度。除测验之外，心理学各领域的观察和结论都要考虑效度。科学家从观测中得出的结论依赖于实验的可靠性和有效性，信度与效度是优秀科学研究不可或缺的成分。

为了发现科学真理，研究者要设计实验并进行理性分析，以减少主观性的解释或个人偏差，同时获得关于人类思想和行为的，可靠且有效的结论。实际上，研究的主题越是离奇，心理学家就越是应该更严格地使用实验技术。

索尔索一直对艺术感兴趣，作为一名认知心理学家，他关注大脑及其工作的机制。他使用已掌握的科学知识和技能科学地研究了自己感兴趣的一个主题：艺术和艺术家。

案例分析

索尔索（2001）对艺术家和非艺术家大脑活动的差异感兴趣。他使用功能性核磁共振（fMRI）对一名艺术家和一名非艺术家进行脑扫描，被试的任务是勾画几何图形和人脸照片。扫描显示，艺术家和非艺术家在人脸加工时大脑特定部位的血流量都有增加。但艺术家的脑血流活动程度没有非艺术家高，这说明艺术家对人脸信息的加工可能更有效率。另外，该大脑部位（右侧顶枕区）只有在被试画人脸时被激活，而在画几何图形时没有被激活。而且艺术家在大脑的右侧额叶区域有更多的活动，索尔索推断：“艺术家通过想象而非肉眼观察肖像作品”。

1.2　科学

如果问学生，“什么是科学？”学生可能会给出许多具体的实例：物理学、化学、生物学等。这种定义表示，科学是描述某些领域的通称。例如，大学生通常要参加一些“科学课程”，而这些课程通常是物理、化学、天文学、植物学、生物学之类的学科。我们虽然能轻易地列举这类例子，但这种定义是不够充分具体的。如果问到为什么化学是科学而历史却不是，为什么物理是科学而音乐却不是，你会发现科学的这种定义就变得有些复杂和令人困惑了。人们通常认为，科学在于事实（可历史也是），或科学在于理论（音乐学也有），或科学使用了实验室进行实验，那天文学或植物分类学又如何呢？它们通常都被认为是科学，但往往不需要进行实验室实验。

科学家和哲学家共同探究科学的适当定义。科学家认为一个领域可以被标识为科学的先决条件包括：收集事实，使用实验作为支持性证据，使用理论作为暂时的解释等。其他定义则强调科学的发展或动态的本质——“科学”是发现新的事实和理论以取代旧的事实和理论，正如爱因斯坦的某些理论取代了牛顿的理论。詹姆斯 · 科南特（James Conant）表述了科学的这一性质，他认为科学是“一系列由实验和观察得出的相互联系的概念或概念图式，并引导进一步的实验和观察”。当一个科学家希望通过实验和观察，弄清楚“什么”与“什么”相关、或“什么原因”导致“什么结果”时，就能很容易地觉察到科学的这种动态过程。新的事实被加入到概念图式中。这些图式（通常称做理论或模型）都是暂时性的，它们在现有的条件下解释了我们所关

注的联系。当新的知识产生，旧的知识和图式就将被新的知识和图式所取代。

当代科学家，可以是物理学家、地质学家、天文学家、人类学家或心理学家，在对其研究领域进行探究时，都遵循着对宇宙结构的一些基本假设。科学思维的普遍假设是，自然是由法则构成并依法则运行的。例如，有个简单但基本的物理法则是，高处掷出的物体会落向地面，就如同铅球从比萨斜塔上坠地一般，而且这种观察能够被重复并得到一样的结果。自然法则是可靠的，反映了宇宙的规则。

正是基于这些由实验和观察得出的假设，我们触及了科学的核心。例如，我们知道了自由落体过程中速度会发生变化，物体在开始时，下落速度很慢，然后会不断加速。通过观察，科学家能够发现这种加速度的基本规律，由此可以引出基本原理或模型，并结合越来越多的宇宙特征，促进理论的不断发展。这一过程显示了科学思维与非科学思维或科学原理和非科学原理的明确区别。因此，科学即你的作为，而非某一对象。

案例分析

古希腊哲学家亚里士多德讨论了自由落体的加速度问题，他认为，按照“逻辑”，重的物体的下落速度应当比轻的物体更快。大石头从屋顶坠地的时间应当比小石子的短，因为大石头更重。许多人仍然相信一颗大炮弹的下落会比一个小铅球快，即使所有其他可能的影响因素都保持一致。“常识逻辑”看上去是有效的，但科学对此存在疑问。伽利略质疑了这一逻辑结论的效度，正如所有小学生都知道的，他决定观察自由落体的相对速度。他的“实验室”是比萨斜塔。

回顾一下伽利略当时的实验，我们可以发现该程序由四个符合当代科学探究标准的阶段组成。

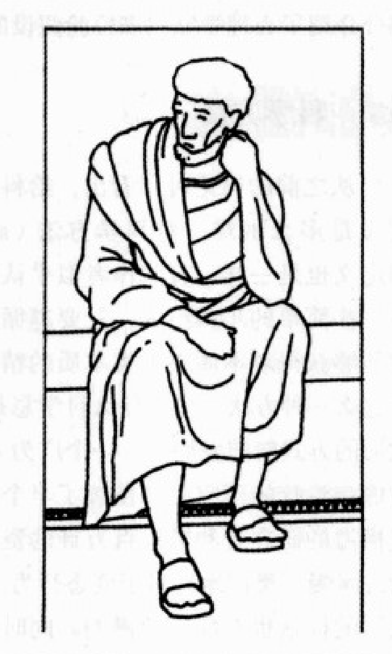
1. **提出假设**。不同重量的物体会以相同的速度落地。
2. **观测**。测量不同重量物体下落的速度。
3. **可重复性**。不同重量的物体可以被多次观测。
4. **引出规律（或模型）**。如果观测结果证实物体重量及其下落速度的某种关系，则可以得出一个普遍性的结论。

当然，这一早期实验存在着某些“控制问题”。这会在后面的内容中详细讨论，现在我们先来了解这个问题。首先，伽利略必须保证两个物体在同一时间开始坠落。例如，如果他让物体从手中坠落，可能会先放开更大更重的那个；又或者他希望支持自己的假设，因而会无意识地先放开更轻的物体。要控制这些问题，伽利略只需设计一个带闸门的箱子，就可以保证物体在同一时间被释放。

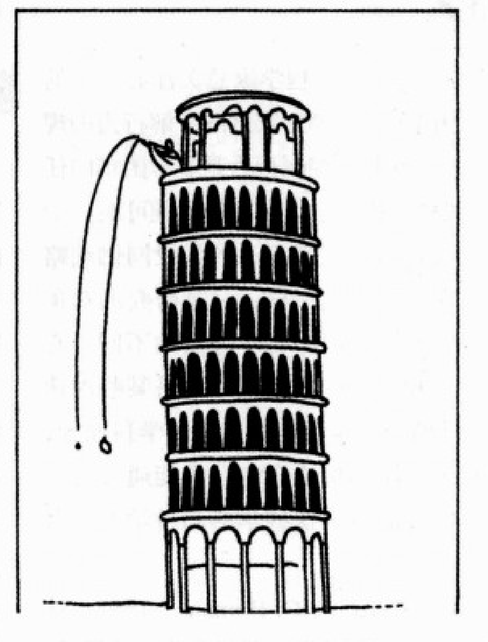
接着是下落速度的判断问题。该实验观测哪个物体率先着地，所以应当采取某种客观指标，以便观察者能可靠地判断物体的着地。

另一个可能影响落体速度的重要变量是大气条件，如空气阻力。观察发现，同样质量的羽毛要比实心球下落得更慢。控制空气阻力变量的一种方法是排出实验室中的所有空气，但由于伽利略的实验室是由比萨斜塔和周围的空间组成，创建这样的一个实验环境显然超出了当时技术的可能（不过，后来有实验观测了真空室的自由落体运动，结果支持了伽俐略的结论）。

自伽利略时代以来，越来越多的精密观察证实了他的简单实验的结论：不论是羽毛或是炮弹，物体在恒定重力条件下以相同速度下落。这一运动定律所依照的原则被称为“等效原理”，被认为是宇宙范围内物体运行的最重要**法则**（law）之一。



亚里士多德：“逻辑表明，重的物体比轻的物体更快落地。”



伽利略：“实验表明，不同重量的物体以相同的速度落地。”

图 1—1　亚里士多德的合理逻辑法和伽利略的实验法

重力定律及其实验很好地解释了两种重要的科学探究水平。第一个水平是基本观察，第二个水平是意识到观察是更大体系的一部分。希望在这本书的学习中，甚至在你作为实验心理学家的职业生涯中，你都能意识到这两种探究水平。例如，在研究的第一阶段，假设你发现了上层社会的孩子上大学的人数比下层社会的更多，这或许是个有意义的观测，但下一步你必须将这一观测融入一个更大的人类行为和社会模型中。如果没有第二阶段的探究，伽利略的观察将是科学史中一个无趣的研究，至多是在某处脚注中会提起有这么一个怪异的意大利人从比萨斜塔上扔下了一些东西。



1.3　事实、理论和推测

首先要弄清楚一些重要的术语。理解事实、理论和推测的区别对于产生研究思路乃至整个科学研究都是非常重要的。**事实**（facts）是可以看到、闻到、听到、尝到或接触到而确定其真实的事物。它们能由人们通过感觉而证实；我们还可以对事实进行测量和操作。**理论**（theory）是对事物本身的可能的解释。你可以收集证据对理论进行检验。有些理论是“好的”，我们甚至把它们视为事实（如重力定律）；有些则需要进一步的科学检验或讨论（如组织干细胞的克隆）；还有些虽然有大量的数据支持，但仍存在社会或政治上的争议（如进化论）。**推测**（speculation）是对未知事物的猜测。推测与理论的区别在于，它既没有足够的数据支持，也无法进行科学的检验。当推测与事实或好的理论不符时，它是有害的，但推测如果被用于激发求知欲或科学探究，则是非常有利的。有些重要的问题目前无法使用科学探究的工具进行研究，这些问题可能难以描述、研究，或进行量化；或许将来可以使用科学工

具更好地研究这类问题。与此同时，请用好奇心来引导你的科学探索。

1.4　心理科学



你可能会问："为什么心理学家总关注心理学的科学性？"答案很简单：心理学家试图了解行为的规律，而精确地了解、预测或控制行为是非常困难的任务。科学探究则提供了执行这类任务的最佳可能。

行为科学家（包括心理学家）的假设同伽利略关于宇宙的法则在本质上是一样；只是研究的对象和使用的技术设备不同，而非科学方法的不同。心理学家致力于思想和行为的研究，而不是某些加速运动的物体；虽然研究对象与专业工具因学科而异，但科学的方法和关于法则构成宇宙，并能通过观察和实验揭示其本质的假设将这些学科和科学家都置于了科学探索的同一道路上。

乔治 · 霍曼斯（George C. Homans，1967）认为，"之所以成为科学乃是因为其目的，而不是结果。如果目的是建立自然特性之间的或多或少的普遍联系，如果对该联系的验证最终归于数据，则这一主题就是科学"。

心理学的科学探究基于与其他科学领域一样的假设：行为和思维的真相可以通过科学分析的方法揭示出来。我们的探究还有进一步的假设：人本质上是一个系统，一个非常复杂的系统，通过科学实验和对实验结果的理性分析能够理解并解释这一系统。有些现象似乎超出了能够进行科学分析的范围，但通常并非如此！由于这些主题的解释通常不稳定，因此实验心理学家们应用尽可能精确的科学方法进行研究，就如应激的皮电记录、阅读障碍儿童的眼动、面孔吸引的研究等。

一般人也了解行为；或者说，他们至少能够与他人共处。每个人都有一些行为的原则，例如"距离产生美"；而另一个原则却说，"眼不见为净。"这种不一致是因为人们对行为的解释总受到个人预期的影响。例如，从基督教原教旨主义者到激进的无神论者，他们对堕胎诊所的描述因其信仰与态度而表现出很大的不同。但心理学家试图避免这种主观主义的影响，他们开展行为的系统研究，形成行为的假设，并通过系统的方法进行检验。在科学方法部分介绍了心理学家用来检验假设的方法与技术。

1.5　科学方法



从之前的讨论可以看出，给科学下一个确切的定义是多么困难；对**科学方法**（scientific method）的定义也是一样。有些作者似乎认为，科学方法是由一些简单的步骤组成，只要遵循那些步骤，就一定能够获得对本质或人类本质的精彩发现。但如果有这么一种方法，那为什么科学总是以缓慢、笨拙、坎坷的方式向前发展呢？一个广为人知的例子就是，对癌症治疗的研究已经持续了半个多世纪，虽有才华横溢的研究者和数以百万计的资金投入，但进展始终缓慢。类似地，对于变态行为（如儿童自闭症）成因的探索也是曲折缓慢的。同时，心理学研究的某些领域则取得了令人瞩目的进展，有些甚至获得了突破性的进展。这里我们只列举当代实验心理学研究中的部分成功例子，如记忆成分的研究、行为的遗传基础、刻板印象对行为表现的影响、阅读障碍的治疗、脑成像研究等。不过，科学进展的缓慢质疑了科学方法是简单的或绝对有效的说法。

心理学研究对科学方法的使用基于两个原则。第一个原则是：科学观察基于感觉经验。我们通过视、听、触、尝、嗅来感觉我们所生活的世界。这种观察通常在某种确定的情境下进行，即实验心理学上被称为的控制条件下的观察，观察结果应当和另一位科学家在类似条件下获得的观察结果一致。这种实验结果可以被重复的特征称为结果的信度，是信度是科学的主要要求。

但是，我们的感觉系统有能力和范围上的局限，正常感觉范围之外的信号往往不被注意；这意味着那些被检测到的信息获得了大于自身的意义。我们把它称做**感觉专制**（tyranny of the senses），表示我们很难意识到在人类的独立知觉范围之外的那些真实现象的重要性。

以无处不在的电磁力为例。在电磁波谱的一端是宇宙射线、γ射线、X射线，而另一端是无线电波、电视波，在中间波段，大约380纳米～760纳米是肉眼能检测到的电磁波。在人类历史的绝大部分时间里，可见光谱范围内的能量才被认为是“真实的”。其他感觉上也存在同样的局限性。即使我们已经意识到无法直接察觉的能量形式是存在的，但仍然会强调那些能够通过普通感觉通道直接检测到的感觉。

技术的发展一定程度上扩展了感觉的范围。许多科学工具和技术能够使独立的感觉系统中“不可见”的东西“可见”。这些工具，如电子显微镜、射电望远镜、光谱仪等将人类正常觉察范围以外的能量转换成人类能够理解的信号。在心理学中，已经有许多精密的仪器（例如用脑成像技术）能让我们看到某个物种的心理，并揭示那些过去无法探知而只能进行猜想或推测的人类和动物的生命秘密。本书中还会遇到许多这类技术，读者们还应注意，如果你感兴趣的技术还没有出现，你甚至可以自己发明一种。

第二个原则是：将感官的观察所得组织成知识的结构。通常实验心理学中的知识结构被称为**模型**（models）。例如，认知心理学家可能会在观察两种记忆及其信息存储、交互规律的基础上提出一个记忆模型。模型来自于观察，并将在足够完整的情况下成为理论。科学方法的一大宗旨就是对理论进行检验或证伪。

科学方法论由各种技术、方法、策略、设计、设备和逻辑法则组成。它们依不同的问题和学科而不同。本书将向学生简要介绍心理学使用的实验技术。基于这一理念，我们先来介绍一种心理学家使用的实验类型，并就此说明实验设计的一些特征。

1.6　实验设计

科学的实验基于有控制的观察。然后根据观察的差异进行推论，这些推论可以用来提出理论和推广到其他类似的情境。**实验设计**（experimental design）是研究者设立的并用于观察比较的前提。通常要求情境中尽可能多的条件保持恒定，这样各组之间的任何变异都被预先进行了确定和控制。实验中被操纵的变量称做**自变量**（independent variable），而被测量的变量称做**因变量**（dependent variable）。对这些变量进行操纵的控制情境和所要测量的效应构成了实验的设计。最简单的设计中只有一个自变量。

我们来看一个“教科书式”的心理学实验设计定义：实验设计即对控制条件的设计，人们可以在控制条件中对人类与动物的活动、思维或行为进行实证观察。实际上，观察是在可验证的假设的指导下进行的，该假设陈述了实验条件及其结果的可能的联系。简单来说，你可以选择一组被试，在一个有控制的环境中引入一个因素，然后测量思维或行为的结果。进一步说明就是，假设你对人瞳孔大小是否影响人脸的“好感度”感兴趣。你可能会假设，当注视人脸的照片时，瞳孔较大的人脸会比瞳孔较小的看起来更具“好感度”（或亲切感）。为了检验这一假设，你希望设计一个实验，或是在一个控制的环境下进行精确的观察。在这个案例中，你需要用**仪器、设备**（apparatus）对思想和活动进行实际测量。本实验可以使用计算机程序呈现不同瞳孔大小的人脸照片，被试根据“好感度”并使用一个等级评定量表评价对每张人脸的印象。在实际实验中，大瞳孔面孔和小瞳孔面孔对两组被试以相反的顺序呈现，这种技术称做反向抵消平衡，这将在后面的内容中进行讨论。在这一案例中，使用反向抵消平衡技术可以控制“好感度”判断受到瞳孔特征而非瞳孔大小的影响。本实验具有几个特征，包括在控制的环境下（实验室计算机呈现）对事件（如瞳孔大小）进行操纵，并进行实证测量（被试判断面孔的好感度）。这个实验中包含了心理实验的所有必要特征。

1.7　科学方法的假设在实验心理学中的应用

科学方法在应用于心理学或其他科学时，有几个关于物理世界和心理世界本质的假设。科学哲学家对此进行了争论，并形成了科学实验的

基础。

秩序。心理和物理世界是有序的，而非随机或偶然的。而且，现象是系统联系的，这些遵循着有序序列的事件能够被观察、解释和预测。例如，我们知道孩子在会走路之前先会爬，对物体的感知先于认识，饥饿的动物比餍足的动物在获取食物奖励时更积极。而性别玩具偏好的研究则发现某些激素（雄性激素）和特定的行为（如玩具选择）有关。

决定论。所有事件都有原因。决定论认为，事件是由可知的并且可检验的因素决定的。女孩的玩具偏好似乎是由较高的雄激素水平引起的（尽管它可能不是这类行为的普遍原因）。心理**决定论**（determinism）假设，心理现象同物理世界中的现象一样，有其原因或前提条件，尽管很多时候难以明确原因到底是什么。在思考心理决定论时要特别注意，一个事件在另一事件之前发生，并不意味着它一定是后一事件的原因。这类结论被称做“迷信解释”，是对心理学思想或活动的错误解释。对心理现象的**迷信解释**（superstitious causes）是指用没有逻辑或经验关联的原因解释思想和行动。心理学家厌恶这种迷信的解释——那看起来就像巫术；然而，科学家和外行都很容易落入这种错误解释心理现象的陷阱。

我们用占星术预测行为的例子来说明这种迷信思维的危害。大家都知道占星术读物和星象图有多么流行。世界上很多媒体都有一个星座运势专栏。在某天的报纸上，星象图给“巨蟹座”的建议是：“避免争执，一个富有经验的长者会给你很好的建议。”给“摩羯座”的预测是：“适合独处，避免和脾气暴躁的人待在一起；一个远方的客户或爱人会给你打电话，并带给你意外的消息。”“太神奇了！”两个预言多少都实现了，只是对其中一人的预言恰好适用于另一个人。“巨蟹座”接到了生活在3 000英里外的一个旧情人的电话，而一个老人则给了“摩羯座”一个如何防止小货车积雪的建议。而其他关于回避争执或脾气乖戾的人的建议都只是些生活常识，并没有更大的意义。当然，仅凭一个例子便拒绝占星师所给的建议同样不是负责任的科学态度，特别是当你对占星术不了解或存在误解的情形下。实验心理学家应当如何检验报纸上那些星座运势的效度呢？

在一次实验心理学课上，索尔索做了这样一个小实验。在两周的时间里，学生每周拿到三次他们的星座运势，然后评价这些预言的效度。结果显示，大多数学生认为那些预言都比较准确。但学生们事先并不知道，只有半个班级的学生拿到了自己的星座运势，而另一半班级拿到的星座运势和他们的生日并不符合。也就是说，他们拿到的是别人的星座运势。然而，在等级评价中没有发现差异。请特别注意：仅因为两组间没有差异，并不意味着在观察中没有发现任何差异。你可以设计更精确的实验检验这个假设。

我们继续心理学中因果关系的讨论。以精神分裂症为例，它的成因可能是儿童期的压力、荷尔蒙失调、遗传因素、疾病或多因素的综合作用；可能还有另外一些没有被发现的因素。最重要的一点在于，精神分裂症行为以及不同类型的玩具偏好、噩梦、尿床、智力的敏锐度、音乐能力、三K党的信仰、好成绩、不良行为等，实际所有的人类思想和活动都不是无中生有的。所有的人类思想和行动都是有原因的。

寻找思想或行为的原因是实验心理学家最为困难的任务，但同时也是令人振奋的，正如伽利略在思考地心引力让不同重量的球以相同速度落地时必然会感到兴奋一般。所有这些探讨的核心在于假设任何思想或行动都存在其原因，这也是心理科学研究的基本假设。

经验论。**经验论**（empiricism）是基于观察和实验去了解某一现象的实践。实证研究依赖于测量（通常是数量化的测量），例如评价学习量、信仰程度、玩玩具的时间等。这些测量以**数据**（data）的表达形式来提供真实的信息。例如，我们用测验分数测量学习过程，用偏好调查测量信仰，用对孩子玩玩具的观察测量玩具偏好。

并非所有实证数据都是有效的。有时研究者会筛选数据以支持预期的观点，或者研究者会忽略研究问题的某些重要方面。第一种情形下，如果可以

筛选支持性的数据并隐瞒对立的数据，研究者几乎可以验证任何假设与理论。不幸的是，那些热情洋溢的政治家们通常会采用这种欺骗方式，他们只采用支持他们的调查数据，而不管别的数据如何。这种实践有时也被称做“怎样用数据说谎”。虽然这种实践对推销候选人有效，但在科学的心理学中没有容身之地。

在第二种忽略相关信息的情形下，事件的原因被忽略了，科学家没有意识到行为的许多可能原因，而只报告了现有数据的部分。这样，研究的结论可能会因为忽视了重要信息而有缺陷。对这一问题的解决需要研究者密切注意科学问题的复杂性并进行更广泛深入的科学探究。

简约性。一般来说，科学家更偏好对自然现象的简单或简约的解释。**简约性法则**（law of parsimony）认为：简单的解释比复杂的解释（更可能正确）更适用。当你说“简单点，笨蛋”时，你可能没有意识到你正在鼓励别人遵循简约性法则！简约的解释要有逻辑性，并要求使用尽可能少的观察或逻辑步骤进行验证。假设你走出你的房子，发现有棵倒下的树挡住了你的行车道。这是一棵很老的树，有些部分已经死了，而当时已经下了整整三天的雨。到底是因为老病的树根部分使树倒下，还是有些游荡的大象趁天黑把它拔了出来？哪种才是最简约的解释？

在推论更广泛的行为本质时，简约性假设很重要；科学家需要从具体发现出发，推论更具普遍性的陈述。在激素和玩具偏好的实验中，数据的表达很简单——雄激素是玩具偏好的前提条件。从这一简约陈述中，我们可以进行有益的推论，例如：男性的女性化行为和女性的男性化行为可能和激素水平有关系。而这一推论的有效性则需要进一步的实验研究和适当的数据收集。

简约性法则还有助于进行同一物种内甚至跨物种观察的普遍性解释。前一种情形可以以人类视觉为例，所有正常人的视觉是类似的。不管你来自马里兰、密歇根或马达加斯加，当你的眼睛和视觉皮层开始加工光线时，你的加工程序都是一样的。尽管人们对看到的事物的解释大相径庭，但人眼在检测光线，将光能转化为神经能量、并将信息上传至视觉皮层时的加工程序对所有人来说都是一致的。简约性法则允许我们进行这样的推论，而不需要对所有人进行检验。另外，法则对跨物种的研究也起作用。例如，非人类（如大鼠、黑猩猩、狗、鸽子等）的视觉特性数据与人类视觉的数据有一些共同的特征；在眼生理的精细水平上，哺乳类对视觉刺激的初始加工是非常类似的。科学家常常为他们的观察寻求简约的解释，以便推论到更大的人类或物种范围。

案例分析

众所周知，男孩比较喜欢拆卸式卡车、拖拉机、赛车和积木之类的玩具，而女孩比较喜欢布娃娃、娃娃家具和厨房设备之类的玩具。是什么造成这种性别差异呢？研究和分析可以如何进行？这个研究的结果能告诉我们性别行为在哪些方面的知识呢？

在医疗机构工作的心理学研究者谢瑞·贝伦鲍姆（Sheri Berenbaum）和梅利莎·海因斯（Melissa Hines）提出了以上这些问题。他们在美国《心理科学》上发表了一篇文章，对玩具的性别特异性偏好的基础进行了检验，文章的标题是“早期荷尔蒙与儿童期性别类型玩具偏好有关”（第 14 章对这篇文章进行了分析）。

从心理学文献和常识可知，男童和女童常会被鼓励去玩某种玩具，而不许玩其他类型的某些玩具。例如，玩布娃娃玩具的男童很快会知道这类行为是不可取的，他们可能会被贴上“娘娘腔”的标签；而不喜欢厨房玩具而喜欢拆卸式卡车的女孩则可能被认为是“假小子”。尽管社会学习和社会压力必然会影响儿童对玩具的选择，但是否可能存在其他的影响因素呢，例如荷尔蒙或遗传？但“常识”是不够的，我们必须进行系统地探究。

为了检验这个观点，贝伦鲍姆和海因斯选择患过先天性肾上腺皮质增生症（CAH）的女孩参加研究，这种病症会产生大量的雄性激素，而通常在男孩体内才会有大量的雄性激素(本研究也考虑了患过同一病症的男孩)。然后,研究者评价并比较患CAH的女孩在玩“男孩的”玩具、“女孩的”玩具和“中性的”玩具时持续的时间。他们发现，患CAH的女孩比来自类似环境，但未患CAH的女孩花更多的时间玩汽车、消防车和林肯原木玩具。作者由此得出结论，“早期的荷尔蒙影响对女性的性别玩具偏好有男性化效应。”这样，本研究就揭示了一个性别行为的新的影响因素。

1.8 心理学实验

我们用心理学**实验**（experiment）这一术语表示至少操纵一个变量来研究因果关系的研究。我们着重强调实验研究，研究者在实验研究中操纵一些因素（变量），控制其他因素，并探究操纵的变量在另一变量上的效应。有些实验中，研究者可能不通过物理方式操纵变量，而通过选择进行操纵，正如先前的玩具选择实验中所使用的方式。研究者没有向女孩的血液中注射雄性激素（尽管这可能是一种验证假设的更为直接的方法，但可能会存在伦理问题，见第8章），而是选择了一些研究者有理由推测其血液中的雄性激素水平较高的女孩。这个案例中，对一个变量的操纵是通过选择来实现的，而非施加因素。这种对某特定事件及其后果（因果）关系的研究正反映了实验研究的特点。

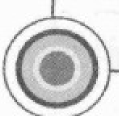
1.9 实验心理学的假设提出

最困扰实验心理学初学者的难题之一就是整理思路并提出一个可检验的假设。由于各种原因，很多有经验的研究者也会面临同样的困难。一个主要的原因是知识的欠缺。新的研究思路很少能从一片空白中自然地迸发出来；实际上，新的思路和假设通常来自个人经验或兴趣，然后建立在已有知识和过去研究的基础之上。因此，忠告就是，让你自己沉浸在心理学某一领域的研究文献当中，并保持你真正的兴趣；阅读、讨论、研究，并精通某些研究领域！知识应当学以致用。当你获取关于某主题、研究前提、结论或技术的知识，要将它们与你其他方面的知识相联系。已经出版的知识并不意味着你不能（或不必）成为细心严谨的知识消费者。心理学新思路的提出与其他学科一样，依赖于对某研究主题的基本元素的掌握以及对所掌握知识的灵活思考；研究者可以将基本想法不断地整合并再整合为新颖而有意义的模式。

新的思路源于旧的思路，新的发现源于旧的发现，而新的假设源于旧的假设。与那些流行小说或电影情节不同，科学的发展通常来自于不断进步的累积，而非单个惊人的发现。当然，我们都渴望那种重大的科学突破，也不应阻止你对获取科学声望的热情。但这种巨大的成就是很少的，大多数研究项目都达不到这样的开创性，但它们仍能为科学知识的整体发展做出有力的贡献。

为了说明这种知识累积的观点，我们来看几年前有一个儿子问某位作者的问题：“是谁发明了汽车？”这位作者尽可能用教育性的方法告诉儿子是卡尔·本茨（Karl Benz）在1886年左右发明了汽车。“哇，他一定是个真正的天才，他得想出发动机、刹车、火花塞，还要让这些一起运转。”这位作者很快发现了一个问题：“我想我可能误导你了。不是某个人发明了汽车的所有部分，也不是某个人发明了电视机、记忆理论或是交响乐。许多人都为汽车的发明做出了重要的贡献。”

心理学知识的发展也是类似的。只要有探究性和创造性的思维、研究的知识和资源、灵活性、奉献精神和坚定的信念，有许多重要的科学真理正等待着未来的科学家去发现。通过归纳和演绎推理，过去的发现产生将来的发现，过去的知识产生将来的知识，而过去的智慧将产生将来的智慧。

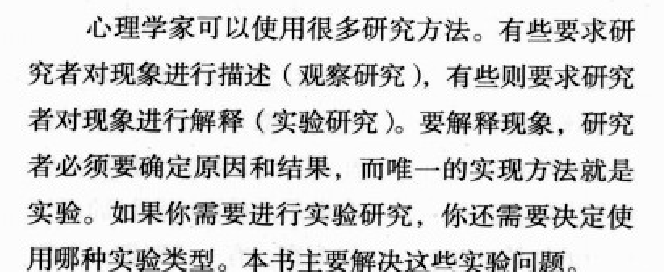


案例分析

基思·雷纳（Keith Rayner）和他的同事对一个转发的电子邮件感到好奇，这个邮件对别人来说不过是继续转发给家人或朋友的普通邮件而已，但他们对此进行了实证性的评价与研究。该邮件宣称：剑桥大学的一个研究项目发现，像“Sentences in whcih lettres weer transpsoed(or jubmled up), as in the setnence you are now raeding”之类的句子，虽然句子中的单词字母的位置被调换了或被打乱了顺序，但人们仍然能轻松地读懂，单词字母的位置对句子的理解并不重要（Rayner, White, Johnson 和 Liversdge, 2006）。他们发现剑桥大学并没有做这样的研究，即表面上看似打乱字母顺序并不会对单词理解产生影响。他们决定科学地检验这一观点。他们让大学生阅读单词字母位置发生调换和没有任何字母调换的句子，字母调换可能发生在单词的开头、中间或词尾部分。结果发现，所有的位置调换都使阅读速度更慢。特别是词头位置的字母调换使阅读速度最慢，接着是词尾位置的字母调换，然后是单词的中间位置的字母调换。看！研究思路的源泉无处不在。

1.10　其他研究方法

心理学家可以使用很多研究方法。有些要求研究者对现象进行描述（观察研究），有些则要求研究者对现象进行解释（实验研究）。要解释现象，研究者必须要确定原因和结果，而唯一的实现方法就是实验。如果你需要进行实验研究，你还需要决定使用哪种实验类型。本书主要解决这些实验问题。



但有许多有趣和重要的心理学问题不需要用实验范式进行研究。其中有些主题可能包括匹兹堡钢铁工人的购物习惯、迈阿密和西雅图双重人格的人数差异或上个世纪的流行趋势等。这些都是很有趣而重要的心理学问题，心理学家可以对此进行科学研究、实证分析并获得可靠的数据。研究者的任务是做出决定并验证这些决定。如果研究者确定了感兴趣的问题，那么第一个决定通常就是他们所要采用的实验或研究类型。最重要的是，你的研究问题决定了研究的方法。因此实验心理学学生要熟悉各种研究方法，才能知道在哪些情况下该使用哪种实验设计。

1.11　观察法



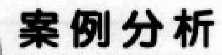
观察法包括系统观察和行为的测量。

行为持续时间和事件频率是观察数据的一般成分。观察法有三种量化行为的方法，分别是频率法、持续时间法和间隔法。

频率法（frequency method）记录特定时间段某特定行为出现的次数。因此，如果你对操场上儿童的攻击性行为感兴趣，你可能会建立一个“攻击性行为”的操作性定义（参见第 3 章），并记录 30 分钟内该行为出现的次数。

持续时间法（duration method）记录特定行为、事件持续的时间。在攻击性行为的例子中，你可以记录每次攻击性行为持续的时间。

时间间隔法（interval method）每隔一段固定的时间观察特定行为是否出现，每次时间间隔为 3 分钟。然后观察者确定该特定的行为（如攻击性行为）是否在该时间段发生。这种信息提供了行为延续性的线索，例如在一个研究中，一只黑猩猩开始某种行为并持续了一系列的时间间隔。



案例分析

内华达大学雷诺分校的艾伦·加德纳（Allen Gardner）和碧翠丝·加德纳（Beatrix Gardner）夫妇开展了一个叫“华舒（Washoe）项目”的灵长类语言和社会化研究。这是最重要、最有意思的研究之一，华舒是一只野外捕获的雌性黑猩猩，大约10个月大时开始和加德纳夫妇一起生活。他们把华舒当做人类孩子一样进行抚养，并以美国手势语言（ASL）作为她的第一语言训练（Gardner 和 Gardner, 1969；Gardner 和 Gardner, 1989）。加德纳夫妇设计了一个词汇测验以证明黑猩猩可以使用ASL和人类交流，他们不仅测试了华舒，还测试了另外三只黑猩猩。他们使用严格的实验控制程序和刺激呈现，并使用经过训练的、独立的观察者来测定这些发现的可信度。物体通过屏幕呈现给黑猩猩，但观察者不能看到所呈现的物体。黑猩猩通过手势语向观察者描述该物体的名称（各位观察者互相看不到对方），并记录下该手势语的名称，然后评估该手势语的准确率（黑猩猩的手势语与真实物体及观察者的理解相比）。结果发现，在71%～88%的时间间隔取样中，黑猩猩都使用手势语正确地描述了物体的名称（与大约6%的机遇水平相比）。

加德纳夫妇可以用手势语和华舒进行一些有意义的信息的交流（比如华舒想要他们帮助她取回一个玩具的时候），他们通过词汇测试在实验室环境下验证了这些现场观察的结果。他们确认，黑猩猩确实能够用手势语告诉观察者那些他们看不见的事物；也就是说，黑猩猩能够使用语言和人类同伴进行自然地交流。

自然观察

顾名思义，**自然观察**（naturalistic observation）是指在自然环境或现场对某些行为或现象进行详尽地观察和记录，有时需要持续很长的时间。一般来说，研究者不操纵环境或对某特定的自变量结果进行测量，而是由社会环境和被试提供行为刺激的数据。在一定意义上，所有人都是自然观察者，也就是说，我们都在某种自然环境中对他人进行观察，比如在机场、超市、酒吧、教室或者影院等。需要强调的一点是，由于被试不是在实验室，而是在自然环境中被观察，所以这其实并不能算是良好的科学观察。

19世纪早期，当查尔斯·达尔文计划去加拉帕戈斯群岛（Galapagos Islands）探险时，他希望能对当地的自然生命进行观察。他并没有设计一系列实验去操纵某些因素并测量其结果。出于达尔文的兴趣和目的，他的自然观察得出了一个科学史上最具影响力的理论。自然观察包含对察觉到的信息的系统记录。观察的环境可以是达尔文时代，不受人类干扰的加拉帕戈斯群岛的一部分，也可以是洛杉矶市中心的一个健身俱乐部；对加拉帕戈斯群岛巨龟复杂的交配仪式和健身俱乐部成员间微妙的调情行为的细致观察都是自然观察。有一段时期，自然观察在美国心理学中似乎成为一种禁忌；但现在这种方法已经获得了更多认可，甚至还一度被认为是收集数据的重要方法之一。即便如此，科学家应当保证自然观察的客观性和系统性，并尽量避免受到个人偏好、感受和偏见的影响。

假设现在你对一个长期困惑人们的主题感兴趣：月圆之夜的行为。很多传说告诉我们（实验心理学家通常会对传说产生质疑），当月亮变圆时，人们会变得不安并且会做一些奇怪的事情；事实上，英文中的“疯子”（lunatic）一词正是来自于其同根词“月亮的”（lunar）。一些不经意的信息有时可以产生一些假设，例如，月圆时人们的睡眠质量会更差，做梦更多，酗酒或更容易狂欢作乐等。警察、

医院员工、消防车司机或其他处理紧急情况的人都证实了月圆之夜更常出现古怪行为，有些研究也有支持该结论的倾向。几年前，我们的一位有进取心的本科生到一个精神病院当晚间护工，他收集了病人夜晚起床次数的数据，并探究与月亮周期的相关关系。结果发现相关是存在的，但其原因并没有真正得到解释。也可能仅仅是因为月光能够提供光线，让病人在上洗手间的过程中不至于扭到脚趾而已。

要收集到月圆水平和夜晚行为的可靠数据，首先就要给我们的研究问题设定一个操作性定义的标准。由于一个人晚间使用洗手间的次数可能只是受到照明条件的影响，所以研究者应当给夜晚行为设定一些更精确的观察标准，例如睡眠过程中被试的姿势、活动的次数和做梦的频数等。所有的观察都应当尽可能隐蔽，以免观察者自身成为额外的干扰因素。另外，所有的因素都被进行了量化，这要用到一些专业设备（虽然许多自然观察是不需要专业设备的）。睡眠者的姿势可以使用睡眠时从一种姿势（如面部朝上）换为另一种姿势（如面部朝下）的次数进行记录；被试的活动可以用安装在床上的四个微动装置精确地测量；做梦频数则可以通过安装在被试眼睑上的微电极对被试的快速眼动（REM）进行记录，而快速眼动通常与梦的活动有关。图 1—2 所示的编码纸可以用来量化这三个变量（研究的因变量）。

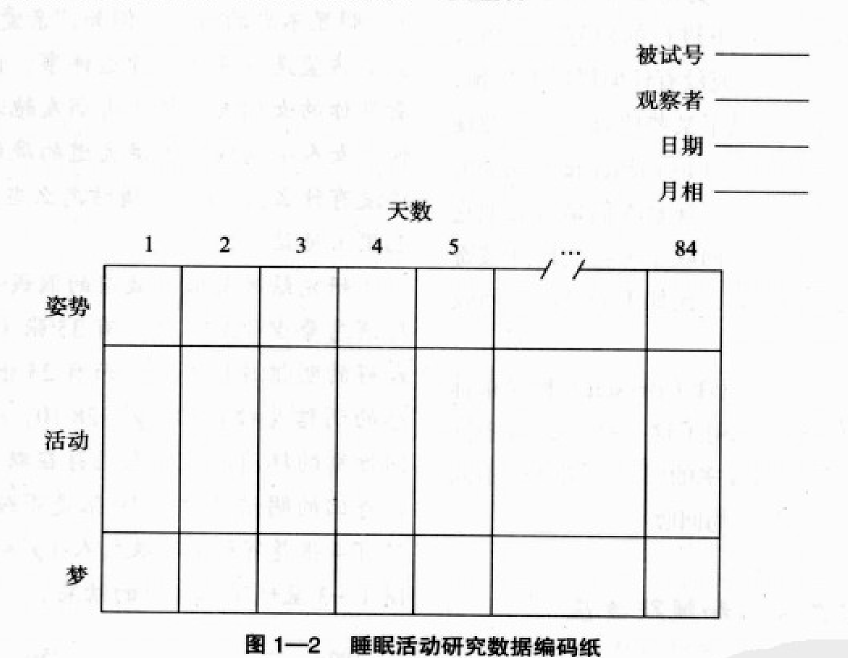


图 1—2　睡眠活动研究数据编码纸

请注意，在表格的右上方有几个空格，那是事后需要填写的信息，包括被试、观察者（可能会有多个观察者收集数据）、日期和月相。如果使用单盲实验设计，则不告知被试实验研究的自变量是月圆水平；如果使用双盲设计，则被试和数据的收集者都不知道自变量。一般认为，双盲控制能够同时避免实验者偏差和被试偏差。

本实验中要详细说明接受观察的被试样本的特征。例如，年长的单身被试的睡眠活动可能比刚结婚的 18 岁小伙子更不容易被察觉；服用了兴奋性精神药物的被试的快速眼动测量可能会比未服药的被试更高；喝咖啡的人的睡眠姿势变化可能会比不喝咖啡的人更多。所有这些被试变量都可能潜在地影响夜晚行为，都应当考虑在内；否则，月圆水平对夜晚行为的影响就会受到其他额外因素的干扰。

许多研究的实证测量都是在自然环境中进行的。下面的案例也是这种类型的研究。

案例分析

迪安（Dean）、威利斯（Willis）和休伊特（Hewitt）1975 年对不同社会地位的人之间的人际距离感兴趣。研究的假设基于一种前提假设：当身份地位较低的人和更高身份地位的人站在一起时，会离得比身份相同的人更远。例如，一个学生和他的教授站在一起时，距离会比和同班同学站在一起更远；而且他会站得离大学院长或校长更远。为了验证这一假设，迪安等人通过隐秘的方式观测了入伍新兵在接近海军军官时站的距离。结果显示，不同身份的人之间的距离会随着身份差异的增大而增大。这一研究揭示了关于个人空间和身份觉察的某些动态特征，它并不适用传统的实验方法，但可以用自然观察的方法进行研究。

现场研究

这种研究的另一例子是**现场研究**（field-based study），这也是在自然环境下进行的研究。它和自然研究的区别在于，自然研究没有任何操纵或干预，而在现场研究中研究者操纵了某些要素。有一种现场研究使用“丢失信件技术”（lost-letter technique），研究者散发一些伪造的信件，观察人们是否会把它们寄还给收信人。研究者对回收率（丢失信件被寄回的数目）进行测量和分析，比如不同社区的回收率差异等。

下面一个研究中，布赖森（Bryson）和汉布林（K. Hamblin）于 1988 年使用了这一技术评价中性消息或坏消息对明信片回收率的影响。请注意不同类别消息和不同性别条件下的回收率。

丢失信件技术和缄默效应

（作者：布赖森和汉布林）

当要向他人传达关于情人不忠的消息时，人们会持怎样的态度呢？本研究采用丢失信件技术的变种——“丢失的明信片”对此进行检验。研究者将贴好邮票、写好地址的明信片放在 180 辆停靠在邮筒附近的汽车的挡风玻璃上，同时还分别附上一张字条，“在你车子旁边发现的，是你的吗？”

1/3（N=60）的卡片上写的是中性的或好的消息（控制条件），例如，“很高兴听说你把事情都解决了。我们都会好起来的。保持联络……”。其中 30 张是寄给男性的，另 30 张寄给女性。

另外 120 张卡片按性别分成相等的两半，内容均是告诉收信人关于他的女朋友（或她的男朋友）明显不忠的消息。例如，“亲爱的鲍勃（朱迪），真不希望是我来告诉你这件事，但我确实在周二看见你的女朋友安娜（男朋友鲍勃）和另一个家伙（女人）从埃尔卡洪大道的廉价酒店出来。也许没有什么，但我不知道该怎么当面告诉你。——巴里（贝丝）”

研究结果与缄默效应的假设一致，被传达的坏消息要少于好消息：有 35 张（58.3%）中性的或好的明信片被寄回，而有 23 张（19.2%）坏消息的明信片被寄回（$\chi^2=28.10$，$p<0.001$）。对不同性别的坏消息收信人还存在双重标准效应：23 张寄回的明信片中有 19 张是寄给男性收信人的，只有 4 张是寄给女性收信人（$\chi^2=12.02, p<0.01$）。图 1—3 呈现了本实验的结果。

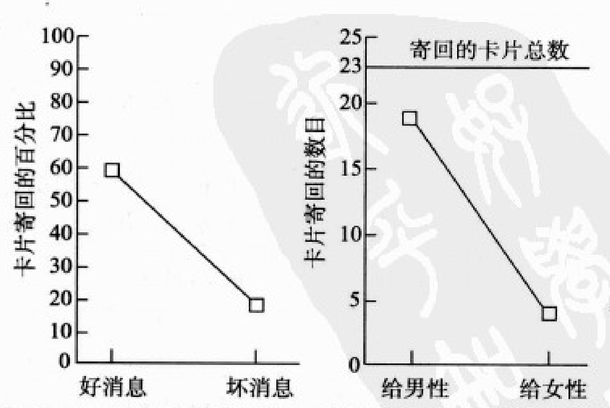


图 1—3　好坏消息卡片寄回的百分比和寄给不同性别的坏消息卡片数目

这些结果表明，人们对传达坏消息普遍存在不情愿的态度（缄默效应），而在愿意传达时也存在双重标准；拿不忠消息的传达来说吧，情人总是最后得知这一消息的，尤其是对女性而言。

为什么人们会倾向于不报告男性的不忠行为呢？有两个看似可能的原因：或许是因为女性的不忠行为更难让人接受，因此人们会倾向于报告；又或许人们认为女性应对伴侣不忠事件的能力比较差，所以应当避免让她们知道真相。

有些现场研究存在严重的伦理问题，因为研究中可能会涉及欺瞒技术的使用，同时又难以获得被试的同意。请记住这一点，当你在学习第 8 章的时候再来回顾这一问题。

1.12　档案研究

我们看到，即使没有对特定的自变量进行操纵，仍然可能进行可靠的、实证的行为观察。**档案研究**（archival research）依赖于现有的记录。这些记录可以是数据、文件或照片的形式，包括你所能搜集到并进行量化的任何已有信息。

一般来说，使用档案研究的观察方式作为因果行为的证据不如实验研究可靠。例如，历史趋势显示沥青的硬度和婴儿的死亡率有很高的**相关**（correlation）。但很难看出这两个变量是如何相互影响的；实际上，出现这种情况很可能是由于第三变量的影响或纯属偶然。所以，对历史趋势的实证观察一定要严格遵循规定的观察标准。我们来看一个历史趋势的实证观察实例。

案例分析

在过去几年，厌食症和贪食症变得越加普遍，尤其是年轻的美国女士。为什么呢？这些不正常的进食习惯是否和理想化的“美女”形象有关呢？如果有关，是否可以对一般文化中的理想化形象的趋势进行研究呢？

马祖尔（Mazur，1986）追溯了女性美的概念变化和某些女性为符合理想化形象而过度适应的历史。该研究项目提出了一些非常特殊的问题，其中一点认为对女性美的理想形象的量化是不够明确的。有些研究者可能会认为女性美是一种可以识别但无法进行操作性定义的概念。然而，马祖尔认为可以通过描述女性形象的一些物理特征来量化女性美的部分属性，就像美国小姐候选者与《花花公子》杂志中的“玩伴”长期以来所描绘的那些特征。该论文中的一些数据见图 1—4。

总体来说，这些数据显示，理想的美女形象正变得更高、更苗条。虽然人们对改变身高办法不多，但可以通过节食调整她们的腰部、臀部和胸部曲线。按照马祖尔对这些数据的解释，有些女性的节食减肥走了极端，并最终导致饮食异常。

我们已经知道，趋势信息的搜集不同于传统实验方法中对某些变量效应的测量。然而，在刚才的案例中，研究者试图学术地研究饮食异常的原因。应当注意的是，饮食异常和其他心理现象一样，通常是由许多因素导致的，理想化的身体曲线可能仅是其中一个。要说明本案例或类似研究中的所有的行为原因，还需要进一步的研究，这时，可能会用到传统的实验技术。我们的结论就是：复杂行为（如饮食异常）通常有复杂的原因，严谨的研究者应当使用各种研究方法探究这些原因。这些方法可能包括传统实验程序和非传统的程序。

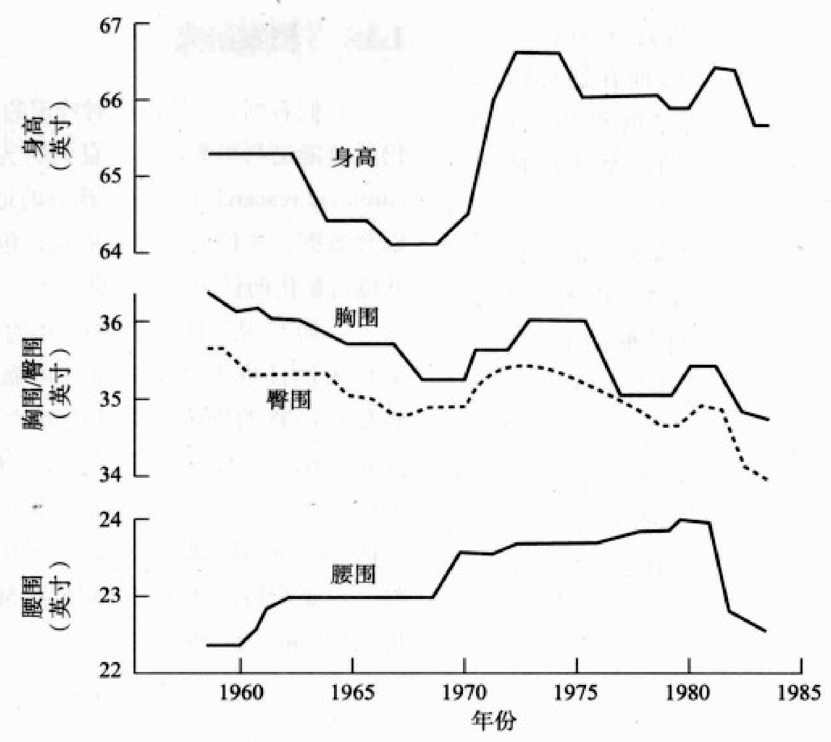
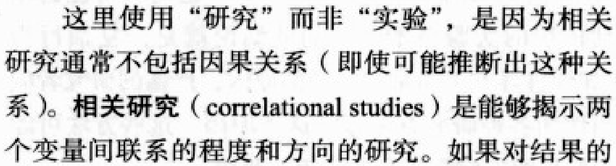


图 1—4 每年“玩伴”中的身体测量数据（平滑后，单位：英寸）

专栏

建立基础率

在基于趋势分析的观察研究中，要建立特定变量的基础率或基线，以进行有意义的比较。例如，有个假设是：婚姻不忠行为会随酒精饮料销售量的增长而增加。要建立这两个变量的因果关系，首先要进行随时间趋势的相关分析。然而，研究者首先要建立外遇和酒类销售随时间变化的基础比率、或是正常发生率。由于这些数据可能不是现有的。研究者有必要收集这些变量的数据，才能进行有意义的观察。



1.13 相关研究

这里使用“研究”而非“实验”，是因为相关研究通常不包括因果关系（即使可能推断出这种关系）。**相关研究**（correlational studies）是能够揭示两个变量间联系的程度和方向的研究。如果对结果的解释没有超出测量的局限范围，那么研究是有价值的；但如果落入**错觉相关**（illusory correlation）的陷阱，对实际上无关的两件事，却误以为存在关系，那么这些研究就是有问题的。

智力就是一种可以在相关研究中进行检验的心理变量，有些人的智商高一些，有些人的智商低一些，而且智商可以被定量地表述。跑步速度是另一个可以量化的变量。我们可以对这两个变量进行相关研究。首先，让一些斯坦福大学的大二学生

进行 50 米赛跑，有些可能跑得很快，有些可能很慢，更多人的速度属于中间水平。接着让这些学生接受一个标准化智力测试，其结果也会出现差异。然后，用一种简单的统计检验方法对这两组数据（智力和跑步速度）进行相关分析并得出一个**相关系数**（correlation coefficient），显示出两个变量之间相关的程度。本研究中可能会出现弱的正相关（即，智力与跑步速度会随着彼此的上下变化而稍微变化）。

案例分析

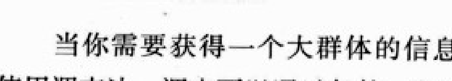
有个相关研究检验了 1970—1980 年旧金山湾地区的酒类消费和出生率。酒类消费和出生率都是变量，妇女出现在两类变量的不同水平上。研究结果发现酒类消费和出生率存在很高的正相关。是否其中一个变量是另一个的原因呢？也许在你的研究中，假设喝酒让更多的孩子出生；但也有另一种可能的解释，就是孩子的出生引起了酒类消费的增长；还有可能这种一致的数据是由于其他变量引起的。仅仅基于相关分析不能进行任何因果关系的陈述。

智力是跑步速度的原因吗（或跑步速度是智力的原因）？喝酒会导致更多的婴儿出生吗（或反过来说）？这些当然都有可能是正确的，但从科学上讲，仅仅建立了两个变量间的相关关系不意味着因果关系的成立。这点非常重要，因为思想家们通常会把两个事件的同时发生看做是因果关系的证据。这种思维通常伴随着一些迷信思想，学生们可以联想到许多佩戴护身符的人，或许他们在某些情况下运气不错，而且那时恰好佩戴了护身符。

另外，相关没有确定因果关系的方向。它们只表示了两个变量间的大于机遇水平的关系。

案例分析

杰西卡·维特（Jessica Witt）和丹尼斯·普罗菲特（Dennis Proffitt）在 2005 年对一种奇怪现象感兴趣：运动员们通常报告，当他们在运动中表现不错的时候，目标看起来特别大（篮筐、高尔夫球洞、棒球或网球）；反过来，当运动员表现糟糕的时候，感觉目标特别小。维特和普罗菲特招募刚刚结束比赛的男女垒球运动员作为被试，让他们指出最符合垒球大小的圆圈。研究者还收集了运动员们在刚结束的比赛中的状态（以便计算击球率）。那么，在运动员的垒球大小知觉和击球率之间是否存在某种关系（相关）呢？确实有！在比赛中击球率高的球员报告的垒球大小比击球率低的那些球员更大。



1.14 调查研究

当你需要获得一个大群体的信息时，就可以使用调查法。调查可以通过邮件、电话、集中调查或当面调查等方法进行。调查看似是一种收集数据的“便捷”方法。但追求便捷并不是选择研究方法的标准！记住，是研究问题决定你的方法。**调查研究**（survey research）如果使用正确，也能够像其他方法一样进行严格的计划和实施。调查问卷要精心构建，调查样本要精心选择，调查过程要适当管理，最后，调查数据要适当分析。如果你只是想知道你

的朋友圈子对下一届总统选举的看法，设计一个调查问卷，然后在他们之间实施就可以了。但如果你想了解大学生对下一届总统选举的看法，就必须细心规划你的取样策略。如果该研究结果要被用来推论到更大总体的看法，调查样本就应当从该总体中随机选取。**样本**（sample）是总体的任何一个子集。**代表性样本**（representative sample）是能够准确代表该总体的样本。**总体**（population）是无法进行测量的、更大的群体，是调查或实验研究数据所推论的被试的集合。样本从总体中抽取，而你最终希望将样本的观察结果推论到整个总体。

1.15　结束语 

其他这些研究方法都可以单独使用，因而也被称做**非实验的方法**（nonexperimental methods）。然而它们通常也被作为实验研究的基本组成部分。你可以开展调查、进行观测、或查阅档案记录，区别在于这时它们是用于实验设计，对变量进行了操作并测量其结果。



第 2 章　心理学文献：理解性阅读并获得研究思路

科学即有系统的知识。

——赫伯特 · 斯宾塞（Herbert Spencer）

近年来，心理学文献有了极大地增长。事实上，论文、书籍和技术报告的数量如此之大，以至于任何研究者都不可能详尽检阅其研究领域的所有资源，更不要说整个心理学领域的知识了。正因为可获得的文献资料众多，而人脑收集和储存信息的能力却是有限的，所以想要成为心理学特定领域的专家似乎是件不可能的事情。然而，这件事乍看起来似乎很难，但本章介绍的一些方法却能使它成为可能。

心理学文献是有组织的知识。由于网络搜索的发展及信息储存的更加便利，心理学文献的组织更加完整，也更容易获得。在了解信息组织的几个原则后，你会发现很容易找到文献阅读的重点，这将极大地有利于你提出自己的心理学研究思路。

2.1　思路、灵感和心理学文献

我们已经讨论过知识来源的问题。坦白地说，要找到一个真正独创性的思路是一个艰巨的任务，对新科学家来讲尤其如此。然而不幸的是，并不存在某种捷径来保证你能够提出新的研究思路；但我们能够给出一些建议，如果再加上你的探究精神，或许就能够让你的实验灵感成为现实。以下就是我们的几点建议。

1. 认真评估你的兴趣所在——思路总是来源于我们的生活经验。
2. 从各种资源中筛选与心理学有关的信息。
3. 有选择地、批判性地阅读心理学文献。
4. 阅读时询问“如果……会……”的问题。
5. 与同事或朋友一起讨论和评估你的构思。
6. 收集该构思的一些相关数据。

首先讨论一下选择性的阅读。我们假设有人正在检索实验心理学的信息。他当然可以无区别地阅读所有能够得到的文献，但即使最贪婪的读者也难以做到这一点。更为有效的方法应当是：选择一个感兴趣的主题，它可以是知觉运动过程、儿童吮吸大拇指的行为、黑猩猩的语言发展、数学学习、工厂领域的人格测试或音乐与疼痛的关系等，然后基于该主题去寻找相关文献。幸运的是，当前的心理学文献都具有高度的组织性，研究者只需要了解组织的系统就可以了。

专栏

心理学资料的其他来源

以下是因特网上心理学相关资料的几种来源，或许有助于你提出新的研究思路并进行**文献综述**（literature review）。

心理科学协会（APS）http://www. psychologicalscience. org

美国心理学协会（APA） http://www. apa. org

心理学网络资源（Psychology Web Resources） http://www. psychwww.com

以下几种专业期刊提供了心理学及相关前沿研究的研究综述。

《心理科学》(*Psychological Science*)，范围包括心理学科下的所有研究领域，特别是具有一般性理论意义和跨领域研究的文章。

《美国心理学家》(*American Psychologist*)，刊登心理学当前论题和心理学各领域的经验性、理论性、实践性研究的论文。

《心理科学最新动态》(*Current Directions in Psychological Science*)，及时提供整个心理科学研究和应用现状的信息，内容简洁易懂。

《公共心理学》(*Psychological Science in Public Interest*)，每一期都包含一个与公共兴趣相关的心理学问题的分析。

2.2 使用因特网

因特网是内容最齐全、信息最丰富的资源。但在使用搜索引擎(如谷歌)和百科网站(如维基百科)进行科学资料检索时要保持谨慎。当然也不是说从搜索引擎上找不到有价值的网址或文献；而是对于刚学习研究文献检索和阅读的学生来说，要从这些信息中鉴别出“好的”和“垃圾”信息有一定困难。除非你的教授让你这么做，否则你最好使用专业的数据库来检索研究论文或课题计划所需的资料。

2.3 电子数据库：索引、摘要和全文

过去想要对某个研究领域进行理解性综述时，研究者需要到图书馆里去翻阅大量的索引、参考书籍和期刊，甚至有可能要跑多个图书馆才能找齐需要的参考资料，这可能要花上研究者几天乃至几个星期的时间。但有了个人电脑、计算机化数据库和因特网，大量的文献检索可能在一个小时之内就能够完成。电子图书馆能够快速、便捷地提供大量信息，而且往往是免费的，研究者有时只需要支付使用网络的费用即可。所有人都能够通过计算机和网络连接登录到这些系统。

和书籍里的索引类似，计算机化数据库也是由心理学论文的标题、作者和摘要（有时是全文）组成。任何人都可以使用一台电脑或终端和网络连接使用电子图书馆。在完成登录后，使用一些简单的指令就能以每秒成千上万次的速度检索想要的资料了。

这里有几种电子数据库可供学生和研究者使用。PsycINFO（参见图 2—1）是 APA 提供的一种电子数据库，提供心理学、社会学、行为科学和健康科学的学术文献的摘要。PsycINFO 包含了 1806 年至今超过 200 万条记录，期刊报告选自于约 2 000 种期刊。章节和书籍涵盖了 1987 年至今出版的世界范围的英文资料，另外还有 1987 年以前出版的大量书籍的记录。现在 APA 仍在以每周更新的形式，每

年添加约 8 100 000 条记录。

PsycINFO 可以让读者从图书馆的服务中获取全文数据。PsycINFO 提供了大量的研究摘要，研究者（从新手到专家）都可以轻松省时地完成大量参考资料的检索。

心理学电子期刊（PsycArticles）是 APA 全文数据库，包含了 APA 期刊和教育出版基金会（Educational Publishing Foundation）出版的期刊的全文。其中收录的文章主要从 1985 年至今，提供 html 和 pdf 两种格式。

High Wire 出版社是由斯坦福大学图书馆提供大量免费的科学文献的全文数据库网站之一，有 918 类期刊和超过 100 万的文献，网址是 http://highwire.stanford.edu。

1

AN: 2005-06794-003
TI: PC_Eyewitness and the Sequential Superiority Effect: Computer-Based Lineup Administration.
PY: 2005
AU: **MacLin,-Otto-H; Zimmerman,-Laura-A; Malpass,-Roy-S**
SO: Law-and-Human-Behavior. Vol 29(3) Jun 2005, 303-321.
PB: Germany: Springer
PT: Journal; Peer-Reviewed-Journal
AB: Computer technology has become an increasingly important tool for conducting eyewitness identifications. In the area of lineup identifications, computerized administration offers several advantages for researchers and law enforcement. PC_Eyewitness is designed specifically to administer lineups. To assess this new lineup technology, two studies were conducted in order to replicate the results of previous studies comparing simultaneous and sequential lineups. One hundred twenty university students participated in each experiment. Experiment 1 used traditional paper-and-pencil lineup administration methods to compare simultaneous to sequential lineups. Experiment 2 used PC_Eyewitness to administer simultaneous and sequential lineups. The results of these studies were compared to the meta-analytic results reported by N. Steblay, J. Dysart, S. Fulero, and R. C. L. Lindsay (2001). No differences were found between paper-and-pencil and PC_Eyewitness lineup administration methods. The core findings of the N. Steblay et al. (2001) meta-analysis were replicated by both administration methods. These results show that computerized lineup administration using PC_Eyewitness is an effective means for gathering eyewitness identification data. (PsycINFO Database Record (c) 2005 APA, all rights reserved) (journal abstract)
DE: *Computer-Software-; *Criminals-; *Face-Perception; *Judgment-; *Witnesses-
DE: Law-Enforcement
PO: Human; Male; Female
AG: Adulthood (18-yrs-and-older)
WEBLH: Find it!
Complete Record | Find Similar | Find Citing Articles
In Database: PsycINFO 2005 Part B.

标题
出版年
作者
来源
出版商
出版类型
摘要
主题词

图 2—1　PsycINFO 记录举例

2.4　心理学文献阅读与理解

检索到感兴趣的领域里的几篇相关文献后，该怎样来消化这些资料呢？许多心理学论文都使用大量复杂的言语、数据和图表来构成技术报告，因此我们需要练习对它们的理解。为了更好地理解文献的特定语言和结构，我们给出两个建议。第一个是阅读论文的策略，第二个是论文阅读记录表。

1. 熟悉心理学文献的一般结构

- 第一是文章标题和作者的联系方式。
- 第二是文章简介或摘要。
- 第三是文章的主体，包括对先前研究的回顾。
- 第四是对实验及其结果的描述。
- 第五是对结果的讨论。
- 第六是文章的参考文献。

初学者应首先浏览全文，确定以上各部分的内容。

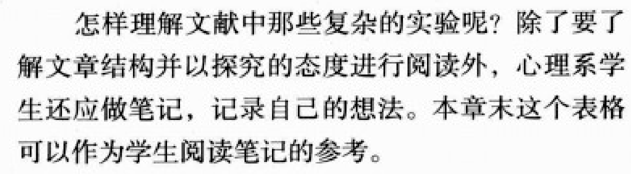
2. 探究式阅读

- 第一步是认真阅读摘要。摘要提供了文章的框架，有助于对文中其他内容的理解。
- 然后，浏览全文，总体把握研究的问题、设计和结论。
- 仔细阅读全文。
- 在阅读过程中提出问题。比如，在文章的开始部分有人可能会问："当前的研究结果和以往的研究结果有什么联系？"有些人可能会质疑研究方法、结果以及研究者得出的结论等。
- 如果你完成了这一批判性的阅读，你可能会发现对填写本章最后的表格有帮助。

3. 形成相关文献检索和阅读的策略

- 第一个策略是阅读该主题的一般性信息，使主题具体化。在教科书里面往往能找到关于某个主题的一般性信息。论著的具体章节和综述性文章［《心理学公报》（*Psychological Bulletin*）或《心理学评论》（*Psychological Review*）等期刊的专业论文］也能提供更多的一般性信息。所有的这些资料，尤其是综述性文章里还有着丰富的参考文献，这对于研究主题的具体化是非常重要的。
- 一篇关键性的文章中提供的参考文献列表是非常好的资源。从中获取并阅读你感兴趣的那些文献，你会了解该作者是如何建构他们的研究逻辑的。
- 文献阅读要先新后旧。如果使用参考文献策略的话，阅读新文献的一个好处就是你可以让那些新文献的作者间接地帮你完成一些文献检索的工作。当然，这时你获得的只是那些研究者认为重要的相关资料，不应该完全替代你的亲自检索与阅读。

2.5　文献阅读记录



怎样理解文献中那些复杂的实验呢？除了要了解文章结构并以探究的态度进行阅读外，心理系学生还应做笔记，记录自己的想法。本章末这个表格可以作为学生阅读笔记的参考。

在阅读完一篇文章之后，你需要完成表格中的各个条目：文章主题、文献来源、研究问题和实施、研究结果和解释、你有哪些疑问以及需要有哪些后续研究来解决产生的问题等。

你可以将这些简要的阅读总结保留下来，并留作进行研究时的参考，有时在后续课程或者毕业设计时仍可以用到。即使资深的研究者也会认为这种形式的记录是相当有用的。如果能够坚持对阅读文献进行总结并记录的话，你可以在很短的时间内就积累起相当多的综述资料，毫无疑问这将有利于提出研究思路。

2.6　结束语

本章主要介绍心理学文献的组织形式、如何检索、如何阅读，以及如何使用这些文献提出新的研究思路。不要害怕文献的收集和阅读，这是你研究工作的开始。下一章我们将分析实验设计的一些基本概念，届时你会对心理学实验有初步的认识。

文献阅读记录表

一般主题

____

文章来源（APA 格式）

____

____

主要内容

____

研究假设

____

____

研究方法

____

自变量

____

因变量

____

研究结果

____

____

解释

____

____

优点

____

____

不足

____

研究思路

____

____

第 3 章　心理学基本实验设计

再多的实验也不能证明我是对的，但只要一个实验就能证明我是错的。

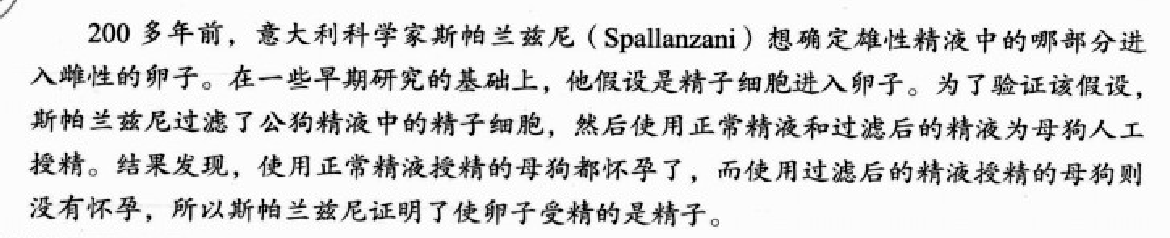
——阿尔伯特・爱因斯坦

3.1　实验心理学的逻辑 

实验心理学其实相当简单。但正如我们在第 1 章中提到的，被试、数据的统计分析、控制观察、理论思考以及研究的内涵等都是科学研究中最为复杂的问题。在涉及这些重要问题之前，我们首先要理解实验科学的简单而重要的方法论。科学方法的一般目的就是理解、描述和解释。

我们通过一个例子来说明实验设计的基本过程。

案例分析



200 多年前，意大利科学家斯帕兰兹尼（Spallanzani）想确定雄性精液中的哪部分进入雌性的卵子。在一些早期研究的基础上，他假设是精子细胞进入卵子。为了验证该假设，斯帕兰兹尼过滤了公狗精液中的精子细胞，然后使用正常精液和过滤后的精液为母狗人工授精。结果发现，使用正常精液授精的母狗都怀孕了，而使用过滤后的精液授精的母狗则没有怀孕，所以斯帕兰兹尼证明了使卵子受精的是精子。

这虽然是一个在 1785 年进行的生物学实验，但它仍然说明了现代心理学和其他科学的实验设计的一个基本原则。首先，斯帕兰兹尼根据已有研究提出一个**假设**（hypothesis），为了验证这个假设，他使用了一种人工授精技术，这样他就可以控制授精的过程以及注射入母狗体内的物质。通过这些控制，他完成了一个两组实验，一组母狗使用过滤出精子的精液授精（实验组），另一组则使用正常精液授精（控制组），这两组在其他条件上均保持一致。由于这两组之间的唯一区别就是在精液中是否存在精子细胞，所以这两组母狗的怀孕比率的任何差异都可以归结于这一操作。正是这一逻辑和实验设计让斯帕兰兹尼得到了有效的结论。

如下框所示，实验过程的基本步骤就是这么简单，事实上很多更复杂的心理学研究也使用这一简单的结构。复杂研究和像斯帕兰兹尼这样的简单实验间的主要区别不在于基本的实验范式，而在于复杂实验有时候存在一些复杂的因素需要良好的控制。也就是说，实验设计其实存在一些基本的原则，在学习和进行实验设计中牢记这些基本原则是很重要的。

实验设计的基本流程

研究假设——→设计实验——→实施实验——→进行观测——→得出结论

3.2 操作性定义

在研究者开始一个实验研究之前，通常要对研究的变量有一个概念上的定义，比如说焦虑、智力、自我卷入水平、驱力、分散练习、强化等。但作为一个可信的研究来说，它不仅仅要能够和读者进行有效的交流，还应该允许其他研究者能够重复它。所以心理学家必须对这些概念进行操作性的定义，即详细定义概念是如何操纵或进行测量的。**操作性定义**（operational definition）是对某概念以具体执行过程来加以定义，比如在研究中是如何测量和呈现的。

操作性定义既要求变量控制应尽可能精确，又希望能尽量保持变量的原意。一方面，像洛奇（Lorge）实验中的练习间隔、摩尔（More）实验中的反馈延迟、贝伦鲍姆和海因斯少女先天性肾上腺皮质增生等都是很容易进行操作性定义的。另一方面，心理学家通常使用智力、敌意、反社会行为或焦虑之类的抽象概念，对这类变量的定义要做到客观而准确是比较困难的。那些抽象的概念到底指的是什么呢？好的实验心理学家会坚持使用操作性定义。这些描述心理学概念的词汇可以和客观环境结合起来。“焦虑”是一个好例子。几乎所有人都知道焦虑是什么，许多词典都比较一致地把紧张的情绪状态作为焦虑的一个主要成分。而在对焦虑这个变量进行操作性定义时，实验心理学家采用了多种方法，如纸笔测验、掌心出汗（PS）技术、皮电反应、心率测试和眼动测试等，这里的每一种操作性定义都可能测量了部分的焦虑。研究者应当能够根据特定的研究问题选择并形成一个适合的操作性定义。



“他提出一个关于可以有多少个天使在一个针尖上跳舞的问题，我认为你们这个问题的可操作程度是不够的，你们说的舞蹈是古典芭蕾、爵士、两步舞，还是乡村摇摆舞……”

专栏

操作性定义练习

以下是心理学研究中常用到的一些抽象概念，请分别建构它们的操作性定义。

焦虑	记忆	色情
创造力	学习	顿悟
攻击性	强化	领导力
智力	自尊	成功
挫折感	态度	死亡
饥饿	惩罚	行为

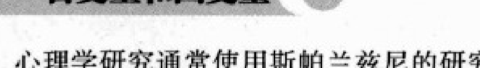
一般有四种对一个概念进行操作性定义的方法。首先是直接询问被试问题（通过调查、问卷或访谈），了解他们是否具有该特质，且程度如何；其次可以测量一些生理指标（如血压等）；也可以测量生化指标（如血检、荷尔蒙水平等）；还可以制定一些行为观察的指标，以便在研究中观察变量的出现情况。当然，有些因素并不能用以上方法进行测量。这里的关键是研究者要认识到定义一个概念的方式是多样的，但必须尽可能选择符合实验目的、客观准确的方式。

例如，假设我们要研究他人的存在对大学生饮酒效果的影响，首先需要操作性定义实验中酒精的类型、量、饮用方法以及饮酒效果的测定方式（恰当考虑伦理因素，参见第8章）。我们将人们带入实验室（有些以10人一组的形式，有些则没有分组），然后按照他们的体重，每100磅提供2盎司的酒。我们可能想知道人们“喝醉”后是什么样的，会做出什么举动。那么该如何定义本研究中的“醉”呢？怎样我们才知道被试是“醉”了呢？以先前的四种操作性定义方法（询问、生理测量、生化测量、观察）为指导，可以轻易地设计出多种方法。我们可以直接询问被试的感受，这可以通过一个或多个口头提问，让他们填写一份评定量表，或者使用其他问卷来完成；如果用生理方法测定的话，可以让他们完成典型的酒后驾车的测试或者评定他们的瞳孔状况；如果用生化测量，则可以测定血液中酒精的含量。最后，如果要用观察法的话，我们需要事先训练一些观察者使用一定的标准（如下表3—1，实验中受伦理因素限制一般不会让被试达到该量表的高分值）来评定他们的行为。每个变量都可以有很多操作性定义的方法，我们可以选择其中一种用于实验。

心理研究中的变量都应当进行操作性定义。

表3—1 杜伯斯基（Dubowski）酒精中毒阶段

血液酒精含量 g/100ml或呼气酒精含量 g/2 101	阶段	临床症状
0.01 ~ 0.05	亚临床（不显）	行为基本正常，稍兴奋、好交际、健谈、自信心加强、防御减弱
0.03 ~ 0.12	欣快	注意力、判断力和控制力减弱，开始出现视-动能力削弱，精细动作测试差，情绪不稳定，关键判断力损失，知觉、记忆和理解受损，感觉反应性下降，反应时间增加
0.09 ~ 0.25	兴奋	视觉敏锐度下降、视-动失调、平衡受损、嗜睡、失去方向感、思维混乱、头昏眼花、情绪夸张、视觉与颜色、形状、运动、尺寸知觉混乱
0.18 ~ 0.30	混乱	痛觉阈限上升、肌肉进一步失调、步履蹒跚、语词含混、情感淡漠、无生气、迟钝、运动功能丧失，对刺激的反应明显减少
0.25 ~ 0.40	昏迷	明显的肌肉失调，不能够站立或行走，呕吐、失禁、完全无意识、反射低迷或丧失，低于正常体温
0.35 ~ 0.50	深度昏迷	没有生机，循环和呼吸能力受损，可能死亡
0.45+	死亡	呼吸停止，死亡



3.3 自变量和因变量

心理学研究通常使用斯帕兰兹尼的研究逻辑。在这个最简单的实验中，研究者操纵一个因素（变量），并保持其他所有因素恒定。

- 被操纵的变量称为**自变量**（independent variable，IV）。
- 被观测的反应变量称为**因变量**（dependent variable，DV）。

斯帕兰兹尼的实验设计的基本原理如下：通常有一组被试接受自变量的处理，而另一组（可以被称做控制组）则没有。除自变量外，这两组被试接受严格一致的处理，在因变量上观测到的任何差异都可以归结为自变量的改变。通过这一控制，研究者可以确定一个变量对于另一个变量的效应，并建立行为之间的因果或其他函数关系。有些人将这种过程称做“控制理论”，这种说法很好地说明了研究的思路。研究的一个关键就是要通过操纵或使之保持恒定的方式对变量进行控制，这样就能更容易地揭示行为的决定因素。

表 3—2 中列举的一些概念与设计或许有利于对心理学实验逻辑的理解。

表 3—2

阶段 1	阶段 2
精液 + 精子细胞（控制条件）	观测结果（怀孕）
精液 - 精子细胞（实验条件）	观测结果（未怀孕）

在一个简单的实验设计中，只对一个变量进行操纵，并观测其在另一个变量上的效应。被操纵的变量就是自变量，而被观测的变量就是因变量。

实验心理学中，因变量通常是对主试问题的直接应答，这种因应反应（dependent response）的使用历史悠久，而且至今仍在使用。心理学中的自变量则有两种：一种是通过研究者系统地操纵一个变量来实现，例如研究者改变对准确反应的动物的奖赏量；第二种是通过对具有某些特质的被试的选择来实现，我们可以依次分析这些不同的特质会对因变量产生什么样的影响。

研究者操纵的自变量

我们通过一个研究来说明因变量与研究者操纵的自变量的使用。研究是关于警察开枪射击黑人和白人嫌疑犯的决策，提出的问题是：警察相对于白人嫌疑犯，是否更倾向于射击黑人嫌疑犯？阿什比 · 普兰特（Ashby Plant）和米歇尔 · 佩鲁陈（Michelle Peruchen）通过计算机模拟的方法来研究这个问题，该研究对自变量和因变量都有清晰的界定。

案例分析

普兰特和佩鲁陈（2005）对警察遇到危险的嫌疑人时，如何迅速判断要不要开枪感兴趣。尤其是，警察的判断是否受到对方种族的影响呢？他们招募了一些警察作为被试进行一个计算机模拟的实验。屏幕上会呈现黑人或白人的面孔，伴随出现的还有一件物品（一支枪或某种中性物品，如手机），警察的反应任务是当面孔伴随着枪支出现时按“A”键——“开枪射击”，面孔伴随着其他物品出现时则按“L”键——“不射击”。结果发现当面孔伴随中性物品出现时，黑人面孔比白人面孔更容易被警察“射击”。也就是说，警察在面对手无寸铁的嫌疑人时，射击黑人的趋向高于白人。不过有意思的是，这种偏差只表现在实验的前半部分。所以研究者认为，警察通过训练是能够克服这种判断上的偏差的。

在普兰特和佩鲁陈的实验中，自变量是面孔图形的种族类别（黑人和白人）以及伴随出现的物品类别（枪支和中性物品）。因变量是“开枪射击”的判断，通过按键盘上的“A”和“L”键来操纵“射击”与“不射击”。下一个实验中，研究者感兴趣的是印象形成的问题。我们对他人的印象形成中是否第一印象要比后续信息更加重要？在阅读这个例子时，你可以尝试去确定该研究的自变量和因变量。

案例分析

阿希（S. Asch, 1952）通过实验来确定第一信息（最先的信息）是否比后续信息（新信息）在形成对某人的印象时作用更大。阿希使用了两组被试，给两组都阅读一系列用于描述某个人的形容词。不过其中一组先接受正面信息而后接受负面信息，另一组的接受顺序则相反。先接受正面信息的这组被试听到的形容词的顺序如下：

聪明的、勤勉的、易冲动的、苛刻的、固执的、嫉妒的。

先接受负面信息的那组被试听到这些形容词的顺序则完全相反。随后阿希让这些被试写下他们对于这个人的总体印象。先接受正面信息的被试把这个人描述为一个有一些缺点但是有能力的人；另一组则把这个人描述为一个由于各种严重的问题而能力受损的人。由于被试接受这几个形容词的顺序的不同，形成了对这个“虚拟中的人”的不同的印象，研究者得出结论认为第一印象在形成对人的观念上是很重要的。

阿希的实验设计

自变量		因变量
形容词的呈现顺序（正向）	——→	对人的印象
形容词的呈现顺序（逆向）	——→	对人的印象

阿希实验中的自变量是形容词的呈现顺序，因变量是被试对“虚拟中的人”的描述。所有其他变量都保持一致，比如两组被试接受的形容词是完全一样的。由于两组处理的唯一区别在于形容词的呈现顺序，阿希认为这正是不同印象的形成的原因，由此可以说明印象形成中首因效应的存在。

研究者选择的自变量

研究者选择的自变量通常指的是**被试变量**（subject variable），如智商、性别、种族、雄性荷尔蒙水平等难以控制或操纵的特征与属性。这些属性都是被试本身具有的，例如一个研究者想要证明权威倾向对概念学习的影响，所以选择了两组研究被试。一组被试在权威倾向标准化测验上的得分高，另一组则在同一测验上的得分低；然后让这两组被试完成同一个概念学习任务，将学习速度或学习量作为研究的因变量。

可以发现在这个研究里，研究者并没有对权威倾向这个概念进行积极地操纵，而是通过选择的方法来获得。一些实验心理学家会将这类使用选择自变量的研究当做非实验研究。严格来说，心理学实验是一种对操纵的自变量结果的观测的研究。但使用被试变量作为自变量同样也属于有效、正确的心理学研究方法。从可操作性来看，心理学中的许多重要问题（如性别、人格类型、诊断分类、种族、年龄、社会地位等类似概念的研究）只能通过使用被试变量加以处理和研究。下面介绍这么一个例子：数学能力是否存在性别差异？

案例分析

本博和斯坦利（Benbow, Stanley, 1980）希望鉴别男孩和女孩的数学能力。他们搜集了9 927名七年级和八年级学生在数学课上的测验分数（两个年级班级数目相等），然后对这

些学生进行学业能力测验（SAT）。在 SAT 的数学推理部分，男孩的平均分要显著高于女孩的平均分。超过 50% 的男孩的得分在 600 分以上（总分为 800 分），而没有一个女孩的得分超过 600 分。研究者还报告了极值分数，男孩的最高分要比女孩高出 190 分。当然这样的结果会引发争议，但研究者支持其数据，并建议科学研究将目标放在寻找造成这种数据分布的原因上。

有许多被试变量都可以作为自变量使用，有个研究让来自富裕和贫困家庭的儿童画一角和 25 美分的硬币，以解释儿童的经济背景和他们对金钱大小估计间的关系。还有许多研究比较男性和女性在许多任务上的反应。还有一个研究比较吸烟人群和不吸烟人群肺癌的发生率。有时这种调查会使用相关分析的方法，比如分析肺癌发生率与香烟使用的关系。但要注意的是，这种方法只是指出了两种现象之间存在关系（如吸烟多与肺癌发生率提高有关系），而不能够表明两者间因果关系的存在。要进行因果关系的推论，还需要有其他方面证据的支持。科学家使用被试变量和相关方法进行的研究是很多的，吸烟与健康危害只是其中的一例。

3.4　实验组和控制组



在很多实验中，各处理组接受自变量的不同水平。有些实验则使用实验组和控制组。虽然在实验中也可以用自变量去描述，但使用实验组和控制组的使用有其独特之处。

实验组（experimental group）是接受自变量处理的被试组，即有研究者的操纵或者自变量的控制。**控制组**（control group）除了不接受实验处理外，其他均与实验组一致。斯帕兰兹尼的实验就是一个很好的例子。接受滤出精子的精液的母狗组成实验组，而接受正常精液的则是控制组。以下例子就使用了实验组和控制组。

案例分析

盲人在行走时善于回避障碍物，但很少人知道其原因。一种假设是盲人发展出一种“面部视觉”，即他们对施加在皮肤表面的空气压力进行反应。另一种假设认为这种回避是使用了听觉线索。素帕（Supa）、科特津（Cotzin）和达伦巴哈（Dallenbach）于 1944 年设计实验对这些理论进行检验。他们让盲人在一个有障碍物（纤维板屏障）的大屋子中行走。研究中有两种实验处理，一种是让盲人戴上面罩和手套（消除“面部视觉”），第二种处理是让被试戴上耳塞（消除听觉线索）。第三种处理是控制组处理，即让被试在未接受任何实验处理的情况下在屋子里正常行走。结果显示，控制组和戴面罩与手套的被试组每次都能够回避障碍物，而戴耳塞的被试组则每次均碰上障碍。根据这些结果，研究者认为盲人知觉并回避障碍物的能力主要是由于听觉线索而非“面部视觉”的作用。

这一研究可以看做是大多数关于正常和盲人被试知觉的研究的简本。在这里要确定自变量可能有些困难，但能够轻易地描述为两个实验处理组（消除面部视觉和消除听觉线索），控制组除了没有接受以上两种实验处理外均与实验组一致。控制组为确定盲人回避障碍的能力是提高还是下降的判断

提供了一个基线水平。研究中的因变量（感觉线索剥夺后的反应能力）则由被试直接走向障碍物的次数来界定。

有时候研究可以需要多个控制组。比如药理研究中就经常使用**安慰剂控制组**（placebo control group）。安慰剂控制组的被试被告知对其施加了能够提高个体行为或治愈某些症状的处理，而实际上并没有对安慰剂控制组施加任何处理。这种安慰剂控制组也常用来检验治疗的效果。以下就是一个使用安慰剂控制组研究的例子。

案例分析

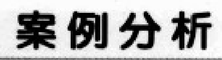
保罗（Paul，1966）通过实验检验两种疗法在治疗演讲恐惧上的效果。他的被试是一所大学里参加公众演讲课程的学生，保罗选择了该课程班里存在严重障碍的67名学生，并分别将他们指派到一种条件下（共四种）。一组为15人，接受行为疗法；第二组为15人，接受认知领悟疗法；第三组为15人，让他们吃无害但也无效的安慰剂药片；第四组为22人，不接受任何的处理，但需要和其他组一样回答研究者的问卷。所有被试需要在治疗开始之前和治疗完成之后各进行一个演说。因变量就用每个人在第二次演说时提高的程度表示，由四位不了解实验分组情况的临床心理学家进行评定。结果显示：行为疗法组的被试100%都取得了进步，而认知领悟疗法、安慰剂组和无处理控制组进步的比例分别为60%、73%和32%。

保罗的实验说明，在有些情况下需要用到不同类型的控制组。如果保罗在他的研究中没有使用安慰剂控制组，对于其研究结果的解释可能就会大不一样。缺少了安慰剂控制组，认知领悟疗法似乎能够有效地改善演讲障碍；反之，我们发现认知疗法作为一种疗法并没有起到应有的作用，至多类似于安慰剂的效果而已，而安慰剂控制组的提高甚至要高于认知领悟疗法组的被试。本实验还说明了一个无处理的控制组的重要性，约有30%的无处理组被试获得了进步，这个指标可以作为评价其他疗法有效性的基线。

不同领域的研究可以使用不同类型的控制组。有个研究把实验组的动物的大脑切除一部分；而控制组的动物也接受了所有的手术过程，只是大脑并没有被切除。这样可以控制并分离可能由于接受手术造成的惊吓对实验组观测结果的影响。要牢记一点，控制组应当除不接受实验处理外，各方面均与实验组保持一致。

保罗的实验还说明了一种用来避免**实验者偏差**（experimenter bias）的重要的控制技术。参与评定被试演说成绩的心理学家都不是在治疗中面对被试的那些人，而且也不知道哪些被试是属于哪个实验组的。这是相当重要的。否则，心理学家们在评定他们在治疗时遇到过的病人的成绩时，可能会出现偏差；而且这四个心理学家会偏向于某一种心理治疗流派，如果知道被试来自于该治疗方法组，就可能会给他们更高的分数；评定者也可能会认为无处理控制组中的被试不会有提高，从而给出相应的评定。保罗选用了独立且“盲”控制的评定者，从而控制了这些可能出现的偏差。

“盲”控制是实验研究中的一种特殊应用。**单盲**（single blind）通常表示实验中的被试不知道自己正在参与实验或不知道所处的实验处理组。**双盲**（double blind）通常使用于药物研究以及需要观察者对实验中的被试行为或提高进行评定的研究，评定者与被试均对实验处理的类型和实验预期结果的类型一无所知。以下一个案例说明了实验心理学研究中实验者偏差可能产生的重大影响。



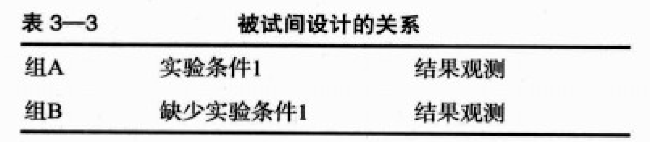
案例分析

罗森塔尔和福德（Rosenthal 和 Fode，1963）揭示了一种明显的实验者偏差。他们选择了一组有实验心理学背景的学生主试，让这些主试对两组大鼠的迷宫成绩进行评定。他们告诉这些主试，其中的一组大鼠属于“迷宫聪明型”(maze-bright)大鼠，另一组则属于“迷宫愚笨型”(maze-dull)大鼠。研究者让这些主试主持一个大鼠走迷宫的研究。结果显示“迷宫聪明型”大鼠的表现显著好于“迷宫愚笨型”的大鼠。不寻常之处在于，其实所有的大鼠都是从同一个标准化样本中随机抽取出来的，根本没有聪明与愚笨之分。难道这里的“迷宫聪明型”大鼠的表现确实好于“迷宫愚笨型”的大鼠？事实可能并非如此，只不过主试在进行观察时，可能产生了一种“聪明”大鼠应当表现更好的期望，而似乎是这种期望污染了观察的结果。

被试内控制

一般来说，心理学实验中有两种简单的控制类型。第一种类型需要有两个或更多的被试组，对其施加不同的条件，其中的一组可以作为控制组，最终对各处理的结果进行比较。这种设计称做**被试间设计**（between-subject design）（参见第 6 章，模型 1 和 2），其主要关系如表 3—3 所示。

表 3—3　被试间设计的关系



组A	实验条件1	结果观测
组B	缺少实验条件1	结果观测

第二种设计类型是**被试内设计**（within-subject design），每个被试都接受两种或两种以上的实验处理。主试在每一个实验处理之后观察其结果，并与其他处理或处理组合后的结果进行比较。其主要关系如图 3—1。

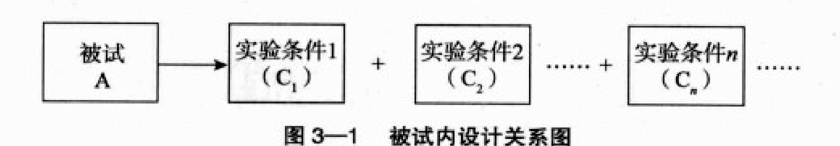


图 3—1　被试内设计关系图

被试内实验设计相对于被试间设计有几个优势。首先需要的被试数量少，每个被试都接受几种实验条件，这和被试间设计每个实验条件需要多组被试相比有优势。其次，由于每个被试进行了被试内的控制，所以不需要对实验组和控制组的被试进行匹配。最后，从统计学的角度来看，被试内实验的变异量（围绕平均分的总体变化）会小于被试间实验的变异量。

不过被试内设计也可能存在一些现实的问题，对某些实验来说并不适用。例如，如果实验中的一种处理会对其他处理的结果产生影响，就不适用被试内设计。如比较两种计算机字处理程序的易学度：Starry Words 程序和 Perfect Words 程序。经典的被试内设计中，一个被试需要在学习完 Starry Words 后再学习 Perfect Words，然后比较学习的容易程度。问题在于，对 Starry Words 的学习经历或许会使 Perfect Words 的学习变得更容易（也可能更困难）。不管怎样，研究者很难知道先前的实验经历对后续实验的影响效果如何，其结果也就不够真实了。

专栏

计算机和实验心理学

几乎在当前的所有实验室中，都能看到计算机的作用。这种作用在未来会变得更加重要。最初，计算机是用于数据的统计分析。心理学家很快就发现了计算机是一种刺激呈现（尤其是计时刺激）的卓越的设备，软件的发展更是使计算机成为实验室中不可或缺的设备。另外，笔记本电脑、PDA、甚至 iPod 手机等都能够用于实验室外的刺激呈现与数据收集。

实验设计的另一个相关问题是**要求特征**（demand characteristics），即被试察觉到研究者的预期而试图做出能够满足这些期待的反应，或者有意图地抑制这类反应。在被试内设计中较容易出现要求特征的问题，因为被试往往接受了多种实验条件，这可能会让被试意识到实验的一些具体内容和目的。于是被试可能会歪曲他们的反应以试图满足（或者挫败）研究者。在实验心理学中要对要求特征有足够的重视，尤其是社会心理学、消费心理学、儿童心理学、认知心理学和变态心理学等研究领域。第4章将会就这些问题提供一些常用的解决方法。

控制的两种方法

先前对控制组和单盲、双盲实验等问题的讨论中已经涉及“控制”这一用语的两种重要的用法。首先，顾名思义，就是研究者控制事件，使之在他希望的时候发生，斯帕兰兹尼对授精过程的控制就是如此。他可以对精液中的物质进行控制，这种操纵或自变量的选择是这类控制的一个范例。另一个范例可以参照研究者是如何观察金丝雀的筑巢行为的。在自然条件下，金丝雀筑巢行为漫长而又分散，要想对之进行实验研究，就需要在研究者希望进行观察的时候让金丝雀筑巢。研究者进行控制的方法就是把金丝雀限制在鸟巢范围内，在需要进行观察时由研究者提供筑巢材料。由于控制了筑巢材料的提供，所以就能够对筑巢行为进行精细的观察了。

第二种控制实验的方法是进行实验条件的设置，以使实验的结果能够用自变量（而非其他变量）的变化来解释，例如保罗通过单盲控制评定者的偏差。另一种保证实验结果的解释性的方法是使用控制组，保罗的实验中就用安慰剂控制组控制安慰剂效应，用无处理控制组控制自然康复效应。这种控制方法将在第5章做更详细的讨论。

3.5　结束语



本章主要讲述实验设计的一些基本原则，许多重要的研究问题都可以由这些基本原则来解释。当然，你可能已经对更复杂、更高级的实验技术感兴趣了，下一章我们将进行这方面的学习。

第 4 章 高级设计技术

不同的实验有不同的用途。

——亚伯拉罕·卡普兰（Abraham Kaplan）

心理学实验研究能让我们得到关于人类与动物心理生活的真实报告，并基于这些报告或事实建构心理学理论或模型。实验心理研究的基本特点在于：心理学实验大多是用来揭示因果关系的，正如食物的剥夺会引发人们的饥饿感，而内心不安则导致焦虑一样。另外有些研究则完全是描述性的，例如个案研究对一个真实环境中的个体进行观察。是啊！生命是异常复杂的，几乎所有的心理现象都不会有唯一的原因，而同一原因往往却能引起我们心理或行为上复杂的结果。正是为了揭示科学领域，尤其是实验心理学领域中的因果关系，研究者们发展出成熟的实验设计与分析的方法。

假设我们要研究人对公路标志的反应能力，不仅要考虑标志的亮度（在完全黑暗的情况下你是看不到它的），还需要考虑其他因素，比如个体对标志含义的理解能力（法语标志或欧洲的信号可能理解起来就存在困难）、个体的意识状态（如果是 24 小时未睡，个体的反应可能比正常时慢）、药物或酒精使用的程度、个体的视锐度以及车辆行驶的速度等。简而言之，单因素设计只是操纵了单一的自变量，并控制其他因素或保持恒定。但生命是如此复杂，行为的原因也是如此复杂。我们需要探究一些略为复杂的，能对多个因素进行处理的实验程序，本章就将分析这些因子设计、准实验设计以及其他复杂的设计。

4.1 因子设计 

单因素实验（即只有一个自变量和一个因变量）是心理学重要的研究工具。但行为显然不仅仅一个自变量的函数。例如，发展心理学的证据表明，如果儿童的父母具有敌意且对儿童过度控制，儿童可能会变得退缩；而如果教养方式有敌意，同时对儿童忽略且不控制儿童的行为的话，儿童可能会形成反社会行为。所以，父母的友善或敌意程度和控制的程度这两个变量都是决定儿童行为的重要因素。

要确定两个或更多自变量（比如你还想把遗传作为第三变量）对一个因变量的影响，研究者就可以使用**因子设计**（factorial design）。在因子设计中，可以同时操纵两个或更多的自变量，并评估自变量在因变量上的效应和自变量之间的交互作用。我们选用比较心理学中的一个案例来说明因子设计的简单应用，研究中研究者分析大鼠在不同的食物剥夺和奖励条件下的奔跑速度。

案例分析

艾伦福罗德（Ehrenfreund）和巴迪亚（Badia）于 1962 年研究了大鼠在改变食物剥夺量和激励条件下的表现。实验使用了一个 5 英尺长的笔直通道，通道的两个端点分别是起始箱和目标箱。因变量测量是大鼠在通道中奔跑的速度。在实验中使用了 20 只大鼠，一

半的大鼠维持其正常喂食量的95%，作为低食物剥夺处理组；而另一半大鼠喂食量为正常的85%，作为高食物剥夺处理组。

实验在目标箱中为高剥夺处理组中的一半大鼠放置了45毫克的食丸（低激励水平），为该组中的另一半大鼠放置260毫克的食丸（高激励水平）。对低剥夺处理组的大鼠也使用了同样的激励处理，这样实验中共有四种实验处理。

1. 高剥夺－高激励
2. 高剥夺－低激励
3. 低剥夺－高激励
4. 低剥夺－低激励

我们来分析艾伦福罗德和巴迪亚的整个实验。四个条件下均有5只大鼠，并测量每只大鼠通过这个5英尺长通道的速度。记录时取用每只大鼠在实验中的最后10次试验的速度（以秒为单位），并计算每只大鼠在这10次试验中的平均速度。表4—1中给出了四个处理组的平均速度。

表4—1 食物剥夺和激励组的平均奔跑时间（秒）

	食物剥夺处理	
	95%（低）	85%（高）
食物激励处理		
45毫克（低）	15.15	13.86
260毫克（高）	13.92	10.26

从表中可知，食物剥夺和激励都能够对大鼠的行为产生影响。高激励组的平均奔跑速度在不同的食物剥夺情况下都要快于低激励组，而高食物剥夺处理组的平均奔跑速度都要快于低食物剥夺组。

这一实验就是一个2×2因子设计。一个变量（食物剥夺量）的两个水平与另一个变量（激励水平）的两个水平两两组合，一共产生了2×2或4种实验处理。第一个变量的每个水平都在另一个变量的不同水平上发生，也就是说，在高激励和低激励条件下均对高剥夺和低剥夺处理的效果进行验证。这种设计还可以加入新的自变量进行扩展，比如在一个2×2×3的设计中，第一个变量有两个水平，第二个变量有两个水平，而第三个变量则有三个水平（例如是大鼠种类：实验鼠、宠物鼠、野鼠）。该设计中则有12种独立的处理。因子设计的一个优势在于，研究者能够弄清不同自变量的组合对于一个因变量的影响。

以下案例是一个关于态度形成研究的2×2因子设计，能够很好地揭示不同变量间的交互作用。

专栏

态度测量

社会心理学家通常要对许多主题进行态度的测量，比如外国援助、太空探索的价值、流产、死刑等。研究者可能要面对几个不同的问题，而其中最困难的是民意调查的问题措辞方式。以1992年总统竞选时进行的一个关于政府福利项目的民意调查为例，一个问题的措辞为："你赞成给穷人增加福利吗？"另一措辞方式为："你赞成为穷人提供更多的帮助吗？"在前一措辞方式下的调查中有19%的人表示赞同，而在后一种措辞方式下则有63%的人表示赞同。

案例分析

有些研究表明，当实验任务是以某种形式诱使人们赞同他们原本反对的事物时，任务的报酬越多，人们改变自身态度的可能性就越大。这和强化理论的解释是一致的，当金钱作为奖励时，报酬越多，人们获得的强化便越大。而另一些研究则发现了完全相反的效应。这些研究支持一种认知不一致理论（dissonance theory）的解释（Festinger，1957）。该解释认为，当人们要对原本反对的事物表示赞同时，就会激发心理的不一致（或是一种不舒适）的状态。而给予的报酬越少，这种不一致的心理就会越强烈，因为少量的金钱并不足以为他们改变自身的态度进行辩护。而为了缓解这种高不一致的心理状态，人们就需要使自己确信自己的态度确实发生了改变。所以在高不一致（低报酬）的条件下，态度的改变会更容易发生。

为了解释这些冲突的研究结果，琳达（Linder）、库珀（Cooper）和琼斯（Jones）认为强化假设只是在被试没有选择的情况下适用，而当被试可以选择是否改变他们对某件事物的态度时，则适用不一致假设。他们的逻辑是，不一致的心理状态仅在人们可以对是否需要做某事进行选择的时候发生，而如果只是被强制去做某事时，是不会产生不一致心理的。

实验采用了 2×2 因子设计。实验任务是由大学生被试每人写一篇文章，支持一个大学生言论自由的限制法案（他们原本是反对该法案的）。研究者要求有些被试必须要写这篇文章（无选择条件），而另外一些被试则可以选择是否要写这篇文章（自由选择条件）；这两个条件下的被试中各有一半被支付了 50 美分的文章报酬，另一半被支付了 2.5 美元的报酬。这样，实验中包含了四种处理，每种处理都随机分派了 10 名被试。研究的因变量是被试对言论自由限制法案的态度改变程度，通过完成一个法案支持程度的量表进行测量。表 4—2 中给出了四种处理下态度的平均改变量，图 4—1 对同一数据进行了直观的描述。

表 4—2　不同可选择性与激励条件下的平均态度改变

	激励水平	
	50美分	2.5美元
可选择性		
无选择处理	0.05	+0.63
自由选择处理	+1.25	−0.07

正的分值表示改变为赞同言论自由限制法案的程度，而负值则反映了被试向原有反对态度的进一步偏移（飞反效应）。这一结果支持了研究者的假设，在无选择处理条件下，2.5 美元的刺激比 50 美分的刺激能引起更多的态度转变，而在自由选择处理条件下则发现相反的结果。

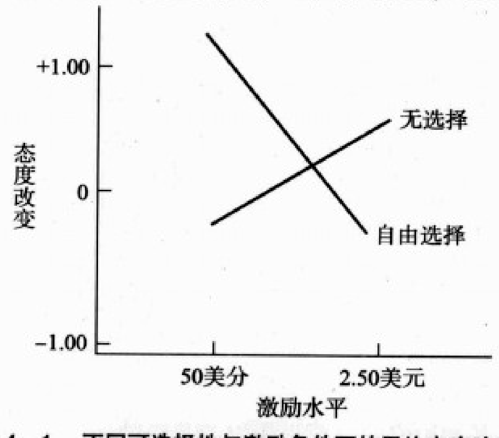


图 4—1　不同可选择性与激励条件下的平均态度改变

方差分析（analysis of variance）适用于对以上研究结果进行分析和检验。这里不去解释方差分析的详细计算（参见附录 A），但了解方差分析的某些逻辑有助于对因子分析逻辑的理解。本实验使用的是 2×2 因子设计，第一个变量的两个水平是“无选择和自由选择”，第二个变量的两个水平是 50 美分的激励和 2.5 美元的激励。在方差分析中，首先对每个变量进行独立分析（可选择性及激励的不同水平），然后再对变量间的交互作用进行分析。当前实验中有两个**因素**（factors），方差分析的步骤包括（1）分析第一个变量（可选择性）的主效应，（2）分析第二个变量（激励）的主效应，（3）分析这两个变量的交互作用。

也可以用图 4—2 的形式来界定 2×2 设计的分析。

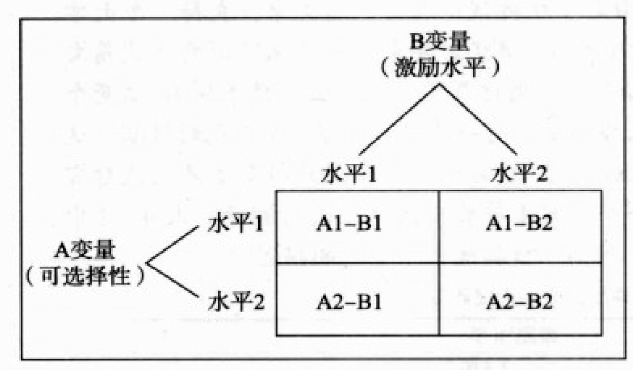


图 4—2　2×2　方差分析设计

变量“可选择性”的主效应可以通过比较无选择处理和自由选择处理下的被试态度改变的平均值进行分析，但不考虑不同激励水平的影响。无选择处理的平均值是把两种激励水平下的态度改变值相加后除以 2（即 – 0.05+0.63=0.58；0.58 ÷ 2=0.29），自由选择处理下得到的平均值则为 0.59。将这两个平均值进行比较，分析变量“可选择性”的主效应。采用同样的方法对变量“激励”的主效应进行分析，结果表明 50 美分激励下态度转变的均值为 0.60，而 2.5 美元激励下为 0.28，将这两个均值进行比较获得不同激励水平的主效应。下一步则需要对两个变量的交互作用进行分析。交互作用的分析是要决定两个变量对因变量的影响是否相互独立。如果一个变量在另一个变量的两个水平上对因变量的影响不同，则说明这两个变量不是独立的。

琳达等人使用这种方法来分析他们的结果。计算结果发现，虽然从均值上看变量“可选择性”和“激励水平”都显示出态度的改变，但并没有达到统计显著。也就是说，这些改变并没有超出均值波动的机遇水平。所谓**机遇**（chance），或称为几率水平，是指研究结果的变异是由于猜测、实验误差，被试未能完美匹配等未控制因素引起的。由于这些未控制因素的存在，需要使用机遇水平这一方法来解释处理无效时的因变量变异。研究中的**交互作用**（interaction）却是统计显著的，在自由选择处理下低激励引起态度的改变，在无选择处理下则高激励引起态度的改变，被试态度的改变是由这两个变量交互决定的。

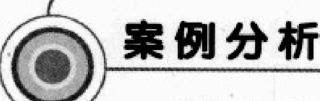
专栏

统计显著

在研究数据收集和分析之前，实验心理学家通常要预期这些数据是怎样进行统计分析的。研究者需要设定一个**显著性水平**（level of significance），即一个事先确定的概率值，依照这个概率值来推断非随机因素的影响。许多心理学实验中的显著性水平表述为“0.05 水平”或“0.01 水平”，这就意味着研究结果基于几率水平发生的可能性只有 5% 或 1%。经过统计分析后符合这些事先设定的标准（可以是 0.05、0.01，甚至是 0.001）的数据就称做统计显著。

应当指出的是，一个统计检验不显著的研究结果仍可能是有意义的。通常在一个试点研究或预备实验中，只使用了少量的被试或观测，其研究结果往往不能达到统计显著水平，但却能告诉研究者是否应该进行下一步的研究，并进行更好的实验控制或更多的实验观测。同时，研究者也要避免使用捕鱼策略，即通过不断操作和调整实验以获得想要的研究结果。

本书中选用的许多统计检验的结果都是“统计检验显著的数据”。在下一个2×2设计的案例中，奇（Chi，1978）使用两类被试和两类任务来研究专业知识对于记忆的重要性。



案例分析

奇使用一个2×2设计研究儿童和成人（A变量）对数字和象棋棋局（B变量）的记忆能力。研究中的儿童都是10岁的熟练棋手，而成年人都是下棋的新手。第一个任务是让被试看一个棋局，当棋局消失后，被试需要重现棋局；在被试任务的数字记忆部分，使用一个类似IQ测验的标准数字序列测量被试的数字记忆。

该设计如图4—3所示。

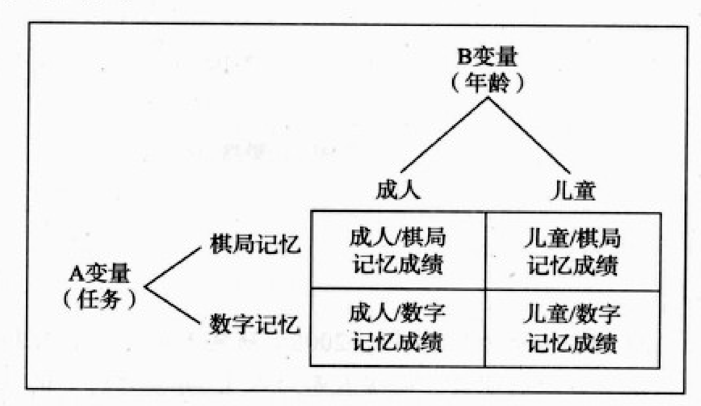


图4—3 成人与儿童对数字和棋局回忆的实验设计

图4—4则呈现了一个显著的交互作用。研究结果显示，专业知识（如象棋知识）会对相应的专业领域的信息记忆有利，但对数字记忆的影响较小。那些作为象棋新手的成人棋子记忆的成绩低于儿童，但在数字记忆中的成绩更好。

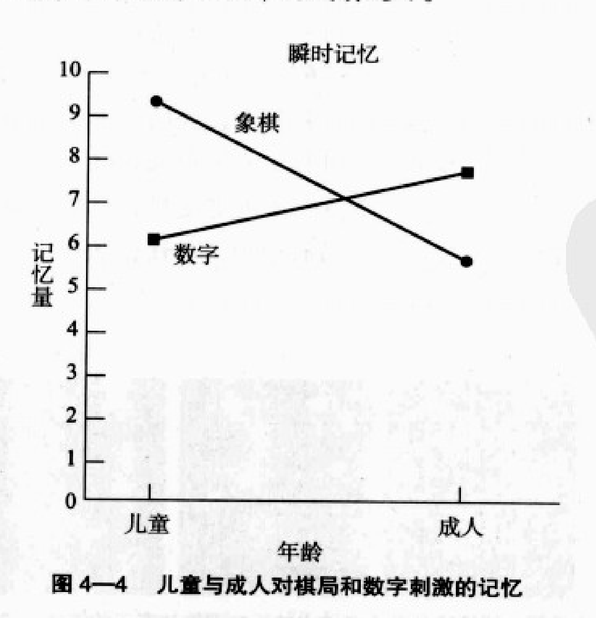


图4—4 儿童与成人对棋局和数字刺激的记忆

2×2设计是因子设计中最简单的一种；随着因子的数量或各因子水平的增加，设计也变得更加复杂。下面是一个社会认知神经科学的研究，虽然只有2个因素，但其中一个因素有2个水平，而另一个则有3个水平。

在这一因子实验设计中，A变量（人脸的种族）有两个水平，而B变量（社会目标）有三个水平。该因子设计如图4—5所示。

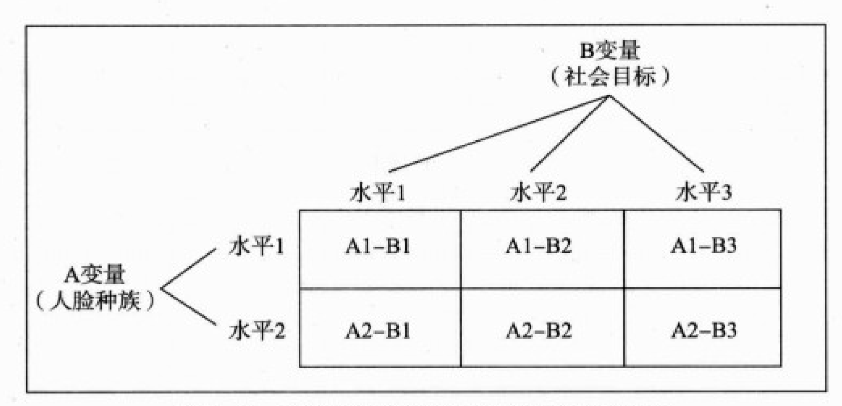


图4—5　社会目标对产生偏见的影响的研究设计

案例分析

惠勒和费斯科（M. Wheeler和S. Fiske，2005）研究不同的社会目标对产生偏见的影响。他们让白人被试看一组人脸照片（有黑人和白人），并依照特定的社会目标（如社会分类、社会个性化或简单的视觉任务）对照片进行反应。所有的任务在一个功能性磁共振成像扫描仪中进行，可以对大脑杏仁核的活动进行观测。已有的研究表明有威胁的反应（潜在危险）能够激活大脑杏仁核的活动。社会分类目标要求被试确定人脸是否已经超过21岁；社会个性化目标则要求被试决定照片中的人是否喜欢一种蔬菜（如芦笋、莴苣或者萝卜）；视觉搜索任务则只要被试判断人脸上是否有一个圆点，被试可以把这些人脸想象成其他任何刺激。被试通过一只手按红色的“no”键和另一只手按绿色的“yes”键，对这三类问题作反应。研究发现在不同的任务间大脑杏仁核的激活并没有差异，任务类型本身不是引起大脑活动差异的原因。但是在不同的社会目标下看不同人脸时，可以发现大脑杏仁核活动的差异。例如，在视觉搜索任务（寻找圆点）下没有发现杏仁核活动的差异；在社会分类任务下则发现黑人面孔导致大脑杏仁核活动的增加；而社会个性化任务下则发现白人面孔导致大脑杏仁核活动的增加（参见图4—6）。

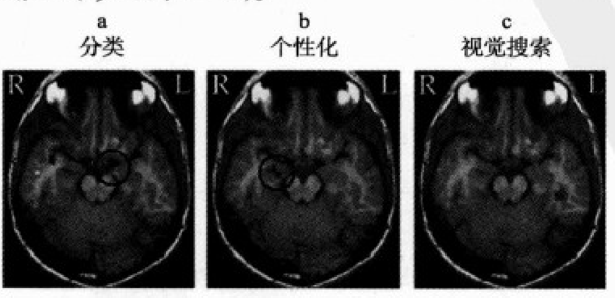


图4—6　研究者使用功能性磁共振成像技术的视觉图像和专门的统计方法确定实验组之间的差异

下面介绍一个索尔索和肖特（Short）的关于颜色匹配的实验，可以用来解释一种更复杂的 3×3×2 因子设计。

复杂的因子设计

下列案例体现了几个实验心理学的重要原则。第一个就是被试内实验控制，即被试同时接受了几个条件的实验处理。第二个是使用反应时的形式对心理过程进行测量，潜在假设是任何心理过程（比如匹配两个信号）都需要一定的时间，可以通过不同实验任务的反应时间的差异揭示许多重要的心理过程。

案例分析

心理学家对信息在记忆中的编码或存储方式很感兴趣，波斯纳（Posner）、艾歇尔曼（Eichelman）、泰勒（Taylor）等人进行了一系列重要的研究，结果发现被试总是先进行字词的视觉编码，而后再进行命名编码。索尔索和肖特（1979）基于以上研究结论进行了一个关于颜色编码的研究。从该实验中我们可以了解到当代实验设计中的几种特征：被试内设计、反应时方法、速示器程序和数据的视觉呈现。

实验呈现给被试一个色块和一个伴随刺激（关联刺激，颜色命名或另一个色块）。伴随刺激可以与第一个色块同时呈现，或者延迟 500 毫秒或 1 500 毫秒呈现。被试判断色块和伴随刺激是否匹配并尽可能快地按键反应。伴随刺激中匹配与不匹配的数量相当，可以保证被试是对匹配任务进行反应。例如在一些试验序列中，被试可能看到一个红色方块，间隔 500 毫秒之后呈现一个词“血”，这就是一个正确的关联刺激，被试要做正确反应。另一个试验序列中，这个被试可能同时看到一个绿色方块和一个“绿”字，这也是正确的匹配。如果在这个被试看到的第三个试验序列中，首先呈现一个蓝色方块，间隔 1 500 毫秒之后呈现一个绿色方块，这种处理就是不匹配的。所以这是一个 3×3×2 的设计，第一个变量是色块与伴随刺激的关系，包括色块 - 关联、色块 - 命名、色块 - 色块三个水平；第二个变量是色块与伴随刺激呈现的时间延迟，有 0 毫秒、500 毫秒和 1 500 毫秒三个水平；第三个变量为色块与伴随刺激是否匹配。

这个设计是一个被试内设计，每个被试都接受每个实验变量的处理。实验结果如图 4—7 所示，被试对色块 - 色块配对（如红色块 - 红色块）的配对反应最快，而对色块 - 关联词的配对（如红色块 - 血）反应最慢。当时间间隔为 1 500 毫秒时，对色块 - 关联词的反应时间也显著缩短，三种匹配处理组的反应时间没有差异。研究者认为这是对颜色平行编码的结果（参见图 4—8）。

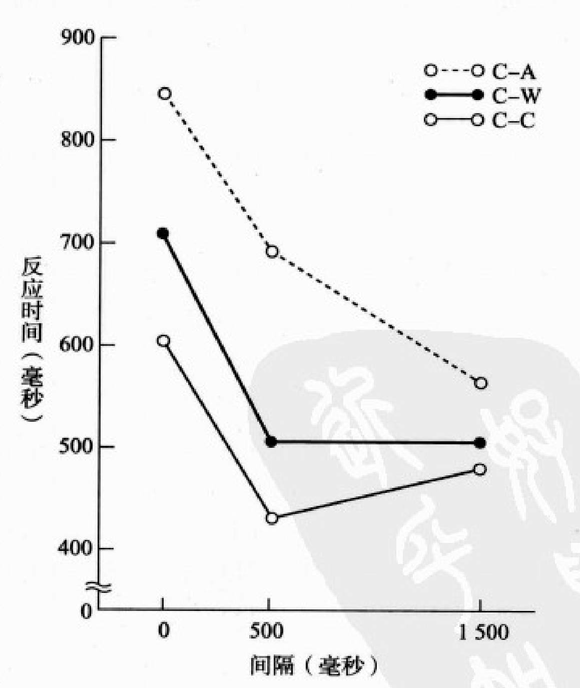


图 4—7 不同配对条件和启动时间间隔下的反应时间：C—A= 色块 - 关联；C—W= 色块 - 命名；C—C= 色块 - 色块

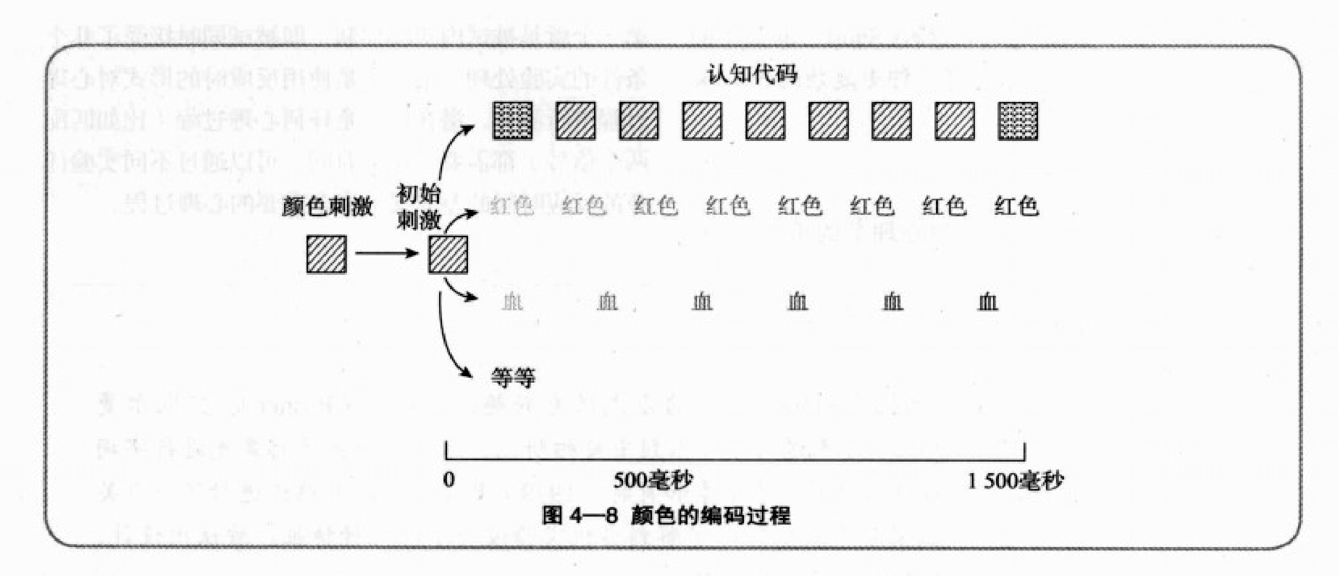


图 4—8　颜色的编码过程

本研究还引出一个主题——**心理测时法**（mental chronometry），即对心理事件速度的测量。这是许多当代认知心理学家善用的一种工具，但其历史却可以追溯到 19 世纪。心理测时法的基本原则是，信息加工需要时间，而复杂的加工过程比简单的加工过程需要更多的时间。

反应时实验通常使用一种迫选程序（forced-choice procedure），要求被试在两个或多个可能中做出选择。由于反应时实验很容易受到练习效应的影响，所以在进行正式实验之前，被试通常要做一系列的准备或练习实验，正如有些研究者说的，"先让他们活动活动食指"。实际上练习或准备实验可不仅仅有这些用途，熟悉实验设备、实验环境、主试、实验程序和实验刺激等都是重要的因素，所以在任何反应时实验或需要注意练习效应的实验研究中都要加以考虑。

这个实验还让我们认识到一种经典的实验装置：**速示器**。这在需要精确呈现短时视觉刺激，并精确测量反应时间的实验中经常使用。经典的速示器是一个三门（或三个视场）的暗箱，能够呈现三种不同的刺激，呈现时间由微小的荧光灯管进行控制。现在，计算机已经可以替代、实现速示器的许多功能。

4.2　准实验设计

坎贝尔（Campbell，1966）和斯坦利用**准实验设计**（quasi-experimental design）的概念来解决许多心理学家在研究中面临的问题。许多心理学家希望研究非实验室情境下发生的行为，准实验设计中的自变量正是来自于这种真实的环境。有时准实验设计也称做事后回溯研究，或是"似乎"设计，因为数据的收集和分析都是在事件发生之后进行的。这些实验设计与自然观察有些类似。准实验设计的逻辑之一是，研究者能够将自变量引入到某种情境中，这种引入方法与情境中产生的该变量是一致的。似乎正是研究者需要对实验变量的引入负责。

为了让读者对这种心理学设计类型的重要性有更好的认识，我们对心理学系研究生部进行了调查，了解最常推荐给准备考试的研究生的书籍有哪些。坎贝尔和斯坦利的《实验和准实验研究设计》（*Experimental and Quasi-Experimental Designs for Research*）是在几个调查中最常被推荐给学生的书籍之一。

在心理学经典实验中，被试的选择要么是随机的，要么是基于某种明确界定的特征（如 19 ~ 25 岁的船员）；然后将被试分派到各实验变量，通常是在严格控制的实验室情境下进行实验；之后再将这一实验的结果推论到更大的总体，或是推论到真实世界中。而真实世界的事件是自然发生的，有时难以选择被试，并将他们放在一个严格控制的实验室环境下收集数据。在实验室怎样研究一场

骚乱呢？假设有人可以做这个研究，这个研究的结果是否可以推广到一场真实的骚乱中呢？这些都是实验和应用心理学家们很感兴趣的话题。实际上，我们对于微观世界的观察（比如在实验室当中）可能并不能有效地代表宏观世界（真实世界）。

实验心理学中的实验室通常是一个高度控制的环境，我们可以称之为**封闭系统**（closed system）。实验室有许多优点，许多心理学实验正需要这种对刺激严格控制的方法。**开放系统**（open system）则指我们很少或没有进行控制的环境，如真实世界。

我们选择一个准实验设计的案例进行分析，该准实验研究了交通事故与交通法规严厉程度的关系。

案例分析

由于过去几年机动车事故死亡人数不断创出新高，康涅狄格州通过了一个限制超速的严厉的法案。在采用了这些措施之后，交通事故死亡人数就下降了。坎贝尔用一个准实验设计对这一现象进行详细的研究。图 4—9 中呈现了康涅狄格州历年的交通事故死亡记录，在通过了限制超速法案之后，1956 年的交通事故死亡人数相对于 1955 年下降了。不过，也许有人会提出很多别的可能的原因，比如可能是路况改善了，天气条件整年都比前一年好，可能有更多的驾驶培训机会等。还有人可能会注意到，图 4—9 中死亡人数记录在 1956 年之后呈持续下降趋势。

还有一种研究方法是将类似的几个州之间的机动车事故死亡人数进行比较。从图 4—10 可知，尽管其他近似的四个州的事故死亡人数在 1951—1959 年也有轻微的下降，但康涅狄格州下降的比例更大，尤其是在 1955 年打击超速之后。这一结果证实了 1955 年所采取的措施对控制交通事故死亡率的效果。

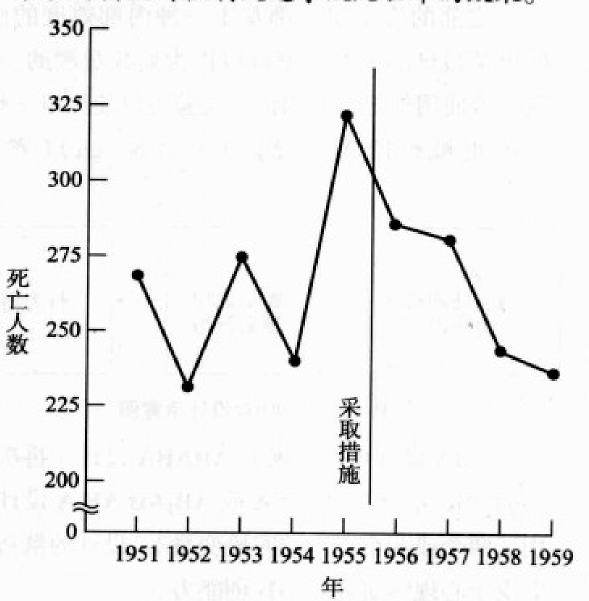


图 4—9 1951—1959 年康涅狄格交通事故死亡人数

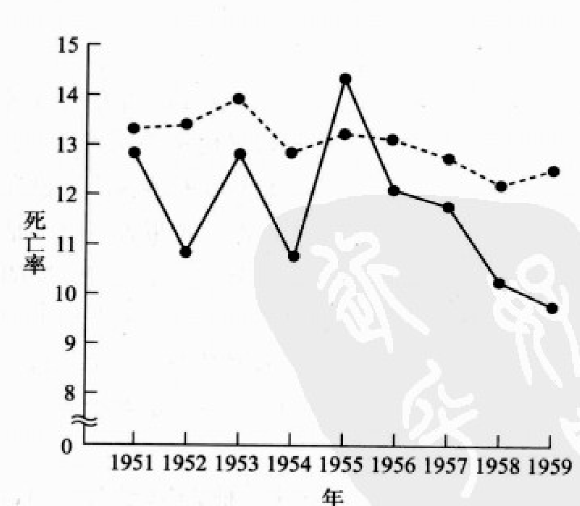


图 4—10 康涅狄格州（下方的线）和其他四州的交通事故死亡率

在这一案例中，严格的实验室实验难以充分验证 1955 年法案是否有效。而这一类问题的重要性不言而喻。虽然没有对这类研究进行实验控制，但准实验设计似乎能够给这类问题提供一种最接近事实的解释。

4.3 功能设计

因子设计通常将几个被试分派到各个实验处理，然后使用相应的统计检验对被试得分的平均数（或百分比）进行比较。这种设计的使用较为普遍，而另一种较常用的研究策略是功能设计。功能设计是"对行为进行实验分析"的方法，是由斯金纳首先提出的，所以该领域的研究者也常被称做是"斯金纳主义者"。**功能设计**（functional design）对实验中的因素采用功能性定义，例如，"惩罚"可以具体定义为一系列决定条件及其行为结果间的联系，定义中的各因素都是可以进行精确测量的。

这一研究方法与本书中其他方法有几点不同。首先，该领域的研究者偏向非理论导向的研究，他们更关注于对控制行为的变量进行检验，而不是对理论进行验证。在功能性设计的潜在假设中，通过对控制行为的变量的系统研究，从研究结果中归纳出理论。

小 *n* 设计

另一点区别在于，功能性设计的研究者有时只用一两个被试进行研究，而其他实验处理方式往往采用大组被试，所以这种设计又被称做**小 *n* 设计**（small *n* design）。自然而然，研究者报告的也不是几个处理组的平均数或方差变异，而是一种典型的反应曲线。这种曲线描述被试在特定实验条件下的典型行为。而且，这种设计有时甚至不使用数据统计分析，而只使用典型曲线来呈现特定刺激条件下的反应模式。

小 *n* 研究中通常使用一种 **ABA 设计**（ABA design）。首先观测被试未接受处理时的行为（A），这一观测有时也称做**基线数据**（baseline data），用以比较并分离实验处理的效应。第二阶段需要引入实验变量并观测其行为结果（B）。最后撤去实验变量，再次观测行为（A）（参见图 4—11）。



图 4—11　ABA 设计示意图

例如，一个行为治疗师可能对一个暴食症病例的处理感兴趣。这里的因变量可以简单地定为来访者的体重（在更精细的设计中，因变量可能包含一系列的生理或心理测量，如新陈代谢率、力气、舒适感、缺失感等）。这样，A 阶段中对未接受处理时的行为的观测就是处理前被试的体重。B 阶段首先引入自变量或实验处理，比如某种行为疗法。这种类型的研究有时会使用两个或两个以上的自变量（例如，治疗师可能会使用积极强化和训练），这时并不能得到清晰的因果关系；但在下结论时，可以分别解释处理条件 1 和处理条件 2 对行为结果的效应。比如在本例中，如果观测到体重的下降，是不能得出行为治疗或训练导致体重下降的结论的，但可以陈述为行为治疗及训练导致体重下降。

之前的许多设计都基于一种内部效度的假设，即单个被试的基线数据可以作为后续观测的一种控制。若使用外部控制的话，实验可以使用以下模式，有时也被称做 **AAA 设计**（AAA design）（参见图 4—12）。

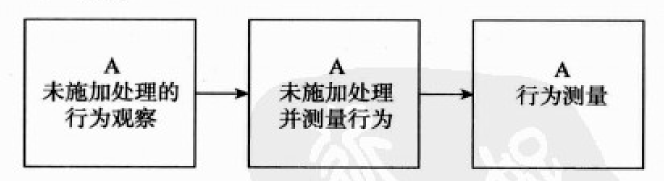


图 4—12　AAA 设计示意图

ABA 设计可以扩展为 ABABA 设计（再次引入实验变量），或 AB_1AB_2A 或 $AB_1AB_2AB_3A$ 设计（使用了两个或两个以上的实验变量）。设计的精巧程度取决于心理学研究者的独创能力。

研究者有时使用高度控制的实验情境来完成这些实验。例如，研究者可能把一只动物放在斯金纳箱里。斯金纳箱通常包括一个按钮（供动物按下，

以获取食物）、食物分配器和一些信号灯。在这一简单的情境中，研究者可以操纵许多自变量，如特定的**强化程序**（schedule of reinforcement）。当依照一种固定比率提供食物强化时，动物需要在按键固定次数（如 1、16、47 或 100 次）后才能获得一颗食丸。依照一种固定时间程序提供食物强化时，动物需要在固定的时间区间内（如每 30 秒或每 4 分钟）至少按键一次才能获得一颗食丸。这些研究中的因变量就是动物按键反应的频率，通常使用累积频率曲线来呈现。以下是福斯特和佩洛特（Ferster 和 Perrott，1968）使用强化程序的一个实验研究。

案例分析

实验使用的仪器是斯金纳箱。箱子面积大约 14 平方英寸，只包括一个小小的树脂玻璃盘子和箱壁上的一个食物盒。盘子与食盒相连，只要推动小盘，食盒中就会掉出一粒食丸。盘子与食盒同时还与一个自动累积记录装置相连，记录器包括滑臂（末端有笔）。滑臂上的笔停留在一个纸带上，随着时间的移动，记录笔会在纸带上沿一个柱面缓慢移动。如果鸽子没有作任何反应，记录笔移动的轨迹几乎是水平的；但每当鸽子啄小盘，记录笔就会在记录纸上沿特定方向移动一小步。最后检查纸带，研究者从特定时间内记录笔向上移动的比率看出鸽子的反应频率。

实验使用的被试是一只有过斯金纳箱实验经历的鸽子。在六周里，鸽子每天要在斯金纳箱里待上一个小时；在整个实验中，鸽子保持在 80% 自然体重的水平。在斯金纳箱中，研究者通过固定比率程序对鸽子啄盘的行为进行强化。固定比率强化有三个水平。第一周，鸽子每啄盘 70 次获得一颗食丸（FR70）；第二周，每啄盘 185 次获得一颗食丸（FR185）；第三周，每啄盘 325 次获得一颗食丸（FR325）。后三周强化水平的顺序与前三周相反。

图 4—13 呈现了在三种不同固定比率程序的强化下，鸽子的行为记录。图中的片断是鸽子日常的典型行为的记录，黑点表示提供了食物强化。各片断的行为可以参照图中右下角提供的一个斜率图进行评估和相互比较。FR70 水平的斜率显示，当鸽子做出反应时，其啄盘的频率大约是每秒 3 ~ 4 次，在这种激励条件下，啄盘的行为几乎是连续不断的，只有在每次食物强化之后才有一个非常短暂的休息。在 FR180 的实验条件下，每次强化之间，这只鸽子休息的时间长的多，但一旦开始，它就以一个非常快的速率啄盘，这一行为会一直延续到下一次食物强化出现。FR325 条件下，鸽子休息的时间更长了，但一旦开始啄盘，其速率与 FR70 和 FR180 条件下的一样快。实验结果表明，获得强化所需要的啄盘数量并不会影响啄盘的动作频率，但影响食物强化与继续啄盘的时间间隔。

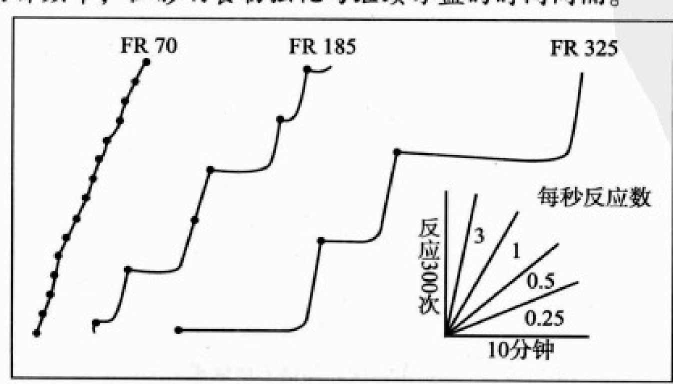


图 4—13　三种固定比率程序下的反应频率

这个例子很好地说明了这种实验设计方法的使用。实验只使用了一个被试，这只鸽子参与了三个条件下的实验（即三种不同的固定比率程序）。实验数据以三种条件下被试行为反应的典型行为曲线呈现，而没有使用高级的统计方法；实验情境本身是高度控制的，强化条件和啄盘反应的记录也要求精确。

这一设计可以应用到许多不同的被试与实验情境。以下是一个控制慢性精神分裂症患者变态行为的实验研究，也使用了同样的设计。

案例分析

持续的强化可能会导致反应强度的下降，这一现象称做“餍足”（satiation），在实验室中，很容易对这一现象进行描述。如果一只动物在很长一段时间不断被施以某种强化，它就会停止被强化的行为。艾伦（Ayllon）使用餍足程序控制一个精神病患者的储物行为。被试是精神病院里的一个47岁的病人，她收集毛巾并储存在自己的房间里。尽管护士们不断把这些毛巾取走，但她收集得更多，每天她房间里收集的毛巾大约在20条上下。艾伦首先设立一个基线，描述在一般情况下被试房间里的毛巾数目。在为期七周的基线观测之后，餍足程序开始进行。护士们不再把被试房间里的毛巾拿走，而是把毛巾带到房间里交给被试，但不作任何解释。护士带来的毛巾从开始的每天7条增加到第三周的每天60条。五周之后，被试已经收藏了625条毛巾，并开始把毛巾拿走，餍足程序终止。

图4—14显示了实验过程中每周被试房间里毛巾的数目。要注意，这并不是一个累积图。在餍足程序之后被试开始将毛巾移出房间，直到第22周，被试房间里的毛巾数目均数降为1.5条。这一平均值持续到第26周。艾伦继续对被试进行了一年的追踪研究，毛巾仍保持在1.5条的水平。被试没有再反复其储藏毛巾的行为，也没有其他替代的问题行为出现。

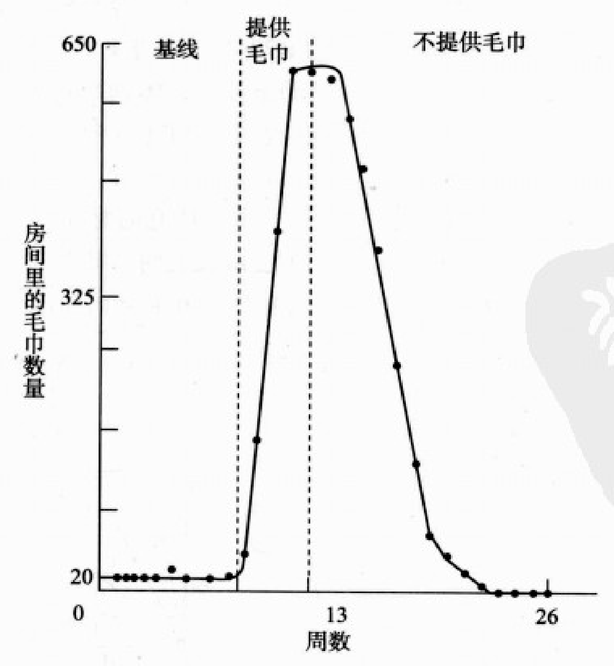
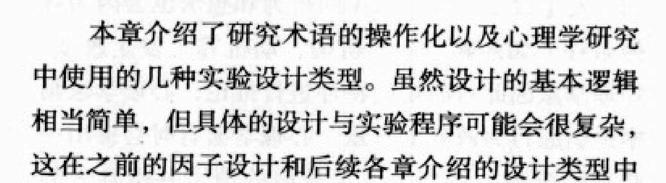


图4—14　处理前、中、后病人房间里的毛巾数量

艾伦的实验是对储物行为治疗的有趣的验证，它还是一个使用功能设计的严格控制的实验。艾伦（1）设立了行为频率的基线，（2）制定一种设计巧妙的实验处理，（3）终止实验处理，（4）并在后期对行为频率进行长期的观测。这里只有一名被试，也没有使用高水平的统计方法。需要强调的是，一个好的实验设计要保证观测到的行为是规律出现的。不论被试是鸽子还是一个精神病人，也不论观测的行为是啄食还是收藏毛巾。

4.4　结果的泛化



本章介绍了研究术语的操作化以及心理学研究中使用的几种实验设计类型。虽然设计的基本逻辑相当简单，但具体的设计与实验程序可能会很复杂，这在之前的因子设计和后续各章介绍的设计类型中可以看到。

需要注意的一点是，研究者通常对他们的实验研究结果和结论能推论到的范围感到困惑。保罗的实验结果显示了行为疗法对认知领悟疗法的优势，但在考虑结果的推广性时，可能存在以下几个问题。

1. 这些结果能推论到多少“病症”上？行为疗法是只对演讲恐惧症还是对所有恐惧症都有优势？这一优势是否仅存在于治疗轻微的心理问题（包括神经症）中，还是在所有的心理问题治疗中都存在？
2. 这些结果能推论到多少“被试”身上？这一发现只适用于大学生吗？同一结果是否同样适用于儿童或中年人？在低智力人群中，是否也能够发现这一结果？
3. 如果治疗时间延长的话，是否还会得出同样的结果？行为治疗作为一种快速疗法是否有效，而领悟疗法则需要更长的时间发生作用？
4. 行为的改善是否可以推论到其他公开讲话的情境？这一治疗的效果是发生在演讲课的情境，所有的课堂情境，还是所有公开讲话的情境？这种提高是暂时的还是永久的？

结果的泛化（generalization of results）是所有研究都要面临的问题。这也反映了单一研究的局限性。在单独一个研究中，研究者总是试图通过严格的变量控制以保证研究结果的有效性。而控制的实现通常是将实验限制于某种特定的行为、某些特定的被试群体、某些特定的观测技术以及特定的时间取样。这些控制技术很好地保证了研究结果的有效，但同时也产生了其结果是否能够推广的问题。这也是为什么一个实验往往引发了其他几个实验，而使科学研究成为一个持续的过程，不断有新的问题值得去研究，新的知识被发现。

4.5　结束语

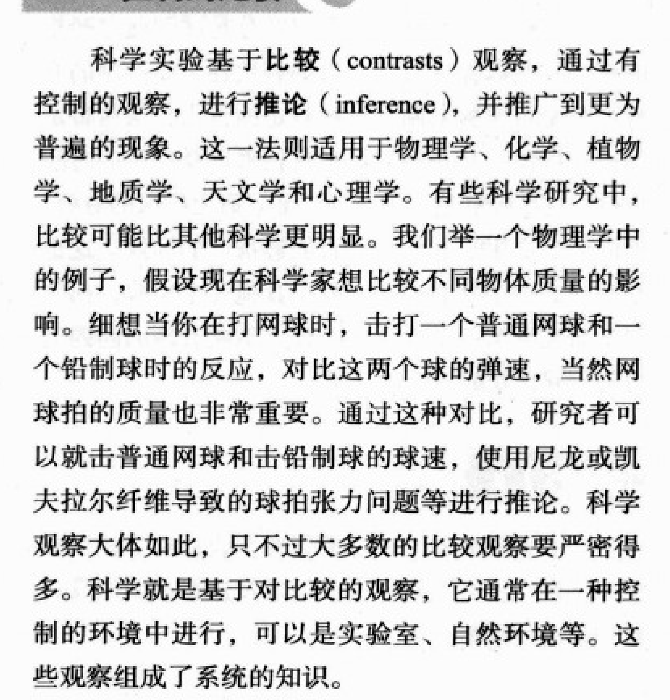
现在你已经对理解实验研究与设计有了相当坚实的基础，下一章将讨论如何通过选择不同的设计对变量进行控制。

第5章　实验设计与控制

某些系统的科学思维方法能够比其他方法带来更大的进步。

——约翰·普拉特（John R. Platt）

5.1　控制的比较



科学实验基于**比较**（contrasts）观察，通过有控制的观察，进行**推论**（inference），并推广到更为普遍的现象。这一法则适用于物理学、化学、植物学、地质学、天文学和心理学。有些科学研究中，比较可能比其他科学更明显。我们举一个物理学中的例子，假设现在科学家想比较不同物体质量的影响。细想当你在打网球时，击打一个普通网球和一个铅制球时的反应，对比这两个球的弹速，当然网球拍的质量也非常重要。通过这种对比，研究者可以就击普通网球和击铅制球的球速，使用尼龙或凯夫拉尔纤维导致的球拍张力问题等进行推论。科学观察大体如此，只不过大多数的比较观察要严密得多。科学就是基于对比较的观察，它通常在一种控制的环境中进行，可以是实验室、自然环境等。这些观察组成了系统的知识。

在心理学研究中，比较通常在几组被试之间进行。一组被试接受一个自变量的几个不同水平，另一组被试则未接受自变量处理。第一组被试称做实验组，第二组是控制组，然后对所有被试组进行比较观察。

5.2　科学推论

科学的理论基于推论。如同铁屑在磁场中的排列会发生改变一样，人类的行为和想法也会因为各种条件（如药物、社会机构、动机等）发生改变。物理学家因此对磁场的特性进行推论，心理学家则推论引起行为改变的变量。心理学家有时会提出一些理论模型，将之推论至类似的情境并指导人的行为。

由于我们观察的效度取决于被试的特征（即自变量的影响），让实验组和控制组尽可能保持一致就显得非常重要。试想，如果一个科学家在磁场中放置了铁屑，而另一个科学家放置的是铅屑，比较观察的结果可能是不同的，科学家可能因此对磁场的特性得出错误的结论。事实上并不是磁力线发生改变，而是放置于其中的物质不同造成的差异。即使观察是有效的，但如果实验缺乏控制，其结果也可能是毫无意义的，甚至可能会误导将来的研究。这一问题对实验心理学家非常重要，也有许多避免因为缺乏良好控制而导致错误结果的方法。下面就将讨论几种可能的错误和控制的方法。

5.3　控制类型

先前的章节讨论了实验设计的策略。在所呈现的案例中，研究者操纵一个特定的变量（自变量）并研究它对因变量的效应。研究者控制自变量呈现的数量、类型和对象，这就是一种实验控制。另外，好的设计要求只有一个变量被作为自变量进行操纵，而其他所有变量在各处理组应保持恒定。**恒定法**

（holding conditions constant）是另一种实验控制。如果各处理组和控制组除自变量外，都接受恒定一致的处理，那么因变量上观测到的任何差异都可以归于自变量。如果存在其他影响结果的变量，就把它称做一个**额外变量**（extraneous variable）或**混淆变量**（confounding variable）。额外的变量意味着这个进入实验的变量是“多余的”。“混淆”的意义则在于，研究者因此难以确认实验结果的原因，是自变量、额外变量，抑或两者皆有？由于结果难以进行推论，研究者必须重复实验，并采用一个能够排除额外变量影响的设计。

只有假设某个变量对因变量有影响，才称之为额外变量。例如，在一个视错觉的实验研究中，诸如被试眼睛的颜色、体重、运动能力或对瑞士奶酪的认识等对被试的知觉都没有影响，研究者可以完全忽略这些被试差异而不加以控制（这些差异通常在实验的代表性样本中被平衡掉了）。另外，视力或错觉刺激的照度等因素可能会影响实验结果，因而需要进行控制。

一个变量是否为额外变量（并需要加以控制）很大程度上取决于研究的因变量。智力对于学习研究可能是一个额外变量，但对于音调差别阈限的实验研究可能就没有影响；对瑞士奶酪的认识和对待流产的态度没有关系，却可能与日用品的感觉识别阈限有关。当然，有人可能会举出一些似是而非的存在联系的证据，例如，感觉阈限的实验也要求被试达到限定的智力水平，对奶酪的知识可能与流产观念存在隐约的联系等。请不要吹毛求疵！心理学领域中有大量重要的问题有待解决，太多的研究新手陷在这些旁枝末节的泥淖里。

好的实验设计需要对真正额外变量的严格控制和对重要内容的细致分析。最精巧的实验也无法挽救一个愚蠢的想法，而一个天才的想法则会因为一个严重的设计问题而浪费掉。良好的知识和判断有助于发现潜在的额外变量，并帮助研究者精确地控制那些可能对因变量产生潜在影响的变量。

本章主要介绍控制额外变量的方法。另一个相关的问题是，如何保证不同的处理组分配到同质的被试，这个问题将在第 6 章进行讨论。

保持其他所有实验条件恒定是控制额外变量效应的一种有效方法。对于额外变量的控制的方法并不多，另一种方法是使用处理组和控制组。所有的实验都使用了第一种控制方法，研究者试图只操纵自变量；有些研究者同时还使用实验组和控制组进一步控制额外变量。在第二种控制方法中，对额外变量和自变量同时进行了实验操纵，这有助于研究者从自变量的效应中分离出额外变量的作用。

保持条件一致

洛奇（1930）的实验比较了集中练习和分散练习的效果。研究的自变量是练习间隔的时间，这是实验唯一操纵的变量。所有其他变量都保持一致：所有处理组的任务一致；所有处理组的练习量一致；所采用的任务能够避免分散练习法的处理组在练习间隔之间进行排练；各处理组的被试能力一致等。这些保持一致的因素都可以是潜在的额外变量。例如，如果某处理组比其他组接受了更多的练习，其结果则可能是由于练习量的差异引起的；如果不同组接受的实验任务不同，结果的差异可能就是由任务变量（比如难度等）引起的；如果某个处理组的被试在任务相关的智力上占优，结果可能就是由于能力水平的差异造成的。通过保持这些因素在各个处理组的一致，洛奇在解释其实验结果的时候就能够排除这些额外变量的影响。这正是控制各处理组除自变量外的因素，并使它们保持一致的基本逻辑。

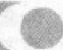
被试变量是个体内在的特性，可以作为研究者选择的自变量。被试变量无法进行随机分派，因此研究者需要严格谨慎地选择被试，以实现对自变量的控制。保持条件一致是好的实验设计的基本保证，心理学初学者都很容易理解这一点，但即便是最有经验的研究者有时也会不自觉地违背这一原则。由于额外变量总是随情境的改变而不同，要列出所有额外变量的清单是不可能的，在有些研究领域中这一问题表现得尤为明显。例如，当实验的自变量是一个被试变量时，它有可能与另一被试变量存在一定的规律性的联系，那么任何实验结果都有可能是不经意间由那个第二变量引起的。这种问题称做“被试变量 - 被试变量混淆”，即可能错误的结论是由于被试变量的混淆或引入导致的另一个原因引起的。

我们以一个研究来说明这种被试变量 - 被试变量的混淆。研究者假设一个专制主义者在学习复杂

材料时要比非专制主义者困难。这一假设的潜在设想是高专制的个体的思考方式比较简单，所以在学习复杂材料时应当比较困难。为了验证这一假设，研究者让高专制和低专制个体学习一些复杂材料。当进行测验时，研究者发现低专制组的记忆成绩显著高于高专制组。几个审稿人对这一研究的质疑之一就是，专制性其实与智力存在某种负相关——专制的个体通常智力表现较差，反之亦然。因此，高专制组学习成绩差可以用智力水平较低来解释，而专制性对这些结果可能没有影响。

如果没有操纵被试变量进行研究，被试变量－被试变量混淆对研究的影响不是很大。但研究者必须注意其他可能的额外变量，有些额外变量及其控制将在本章中进行介绍。

5.4 实验范式



实验心理学中的**范式**（paradigm）是指研究者组织其研究所使用的模型或模式。在这一部分，我们介绍以下四种实验范式及其控制方法。

模型 1：被试间设计

模型 2：匹配组被试设计

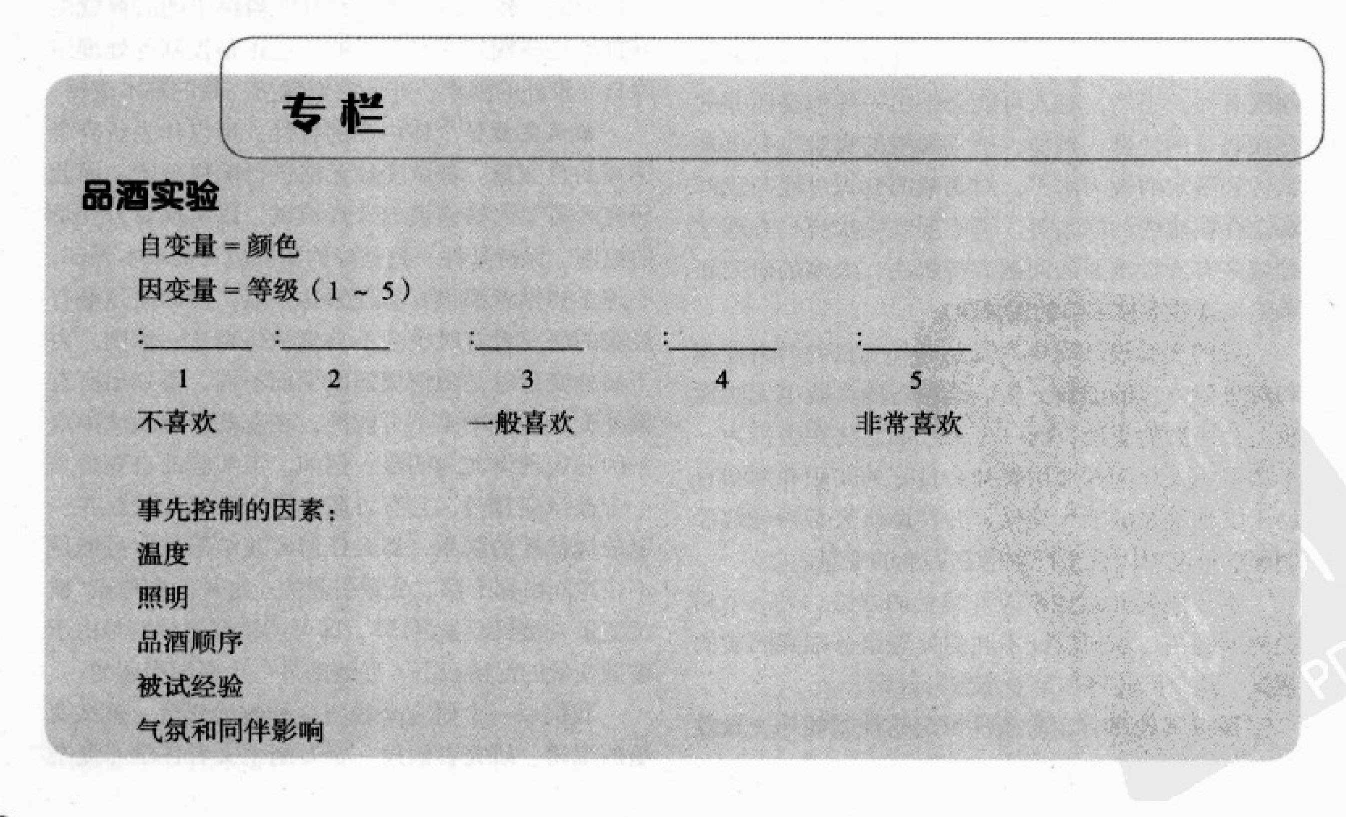
模型 3：被试内设计

模型 4：因子设计

这些范式和控制的一些内容将在后续部分进行介绍。这里，我们通过一个简单的例子来了解四种设计范式。

假设研究者对酒的颜色与个体的饮酒量的关系感兴趣。他们设计出一种装置，能够改变酒的颜色，但不改变酒的口感。进行测试时，某种酒的颜色本来是暗红色，它可以变为深绿色。实验的因变量是使用一个五点量表对喜好程度进行评定；自变量是酒的颜色。由于“喜好程度”不是一个精确的术语，所以需要事先对其进行操作性定义，界定一些重要的因素。比如，喜好程度可以包括口感、香味和颜色等。

即使这样一个简单的研究仍可能存在一些实验设计的问题，下框中列举的问题包括：酒的温度（应当保持恒定）、照明（由于需要看酒的颜色，不应太暗或太亮）、品酒顺序（先品尝的酒可能会对后面的有影响）、品酒人的经验（老练的品酒人和只会牛饮的品酒人的评价可能不同）、周围的气氛（测试应当独立进行，因为同伴压力可能会对评价有影响）。这个实验可以有几种设计方法。



专栏

品酒实验

自变量＝颜色

因变量＝等级（1 ~ 5）

:________ :________ :________ :________ :________

1 2 3 4 5

不喜欢 一般喜欢 非常喜欢

事先控制的因素：

温度

照明

品酒顺序

被试经验

气氛和同伴影响



模型 1：被试间设计

被试间设计也称做**随机化被试设计**（randomized subjects design）（模型 1）。在这种设计中，一组被试接受的是加工后变色的酒（实验组），另一组被试接受原色酒（控制组）。两组被试是相互独立的。实验的被试样本为 16 名 21 ~ 30 岁的大学生。尽管没有明文规定样本的大小，但每组 8 个被试（或观察）应当是这类实验的最低限制；如果样本能增加一倍会显得更严密一些。样本中的被试被随机分派给实验组或控制组。为方便起见，我们将被试标志为 S_1、S_2、S_3……S_{16}。被试分派程序如下表所示。

表 5—1　　模型 1：被试间设计

	实验组（人工调色后的酒）数据		控制组（原色酒）数据
S_1	1	S_9	5
S_2	2	S_{10}	4
S_3	2	S_{11}	3
S_4	1	S_{12}	5
S_5	2	S_{13}	5
S_6	1	S_{14}	4
S_7	1	S_{15}	4
S_8	4	S_{16}	5

品酒分组：自变量——酒的颜色
因变量——质量评价

在实验中，我们得到了一些类似于真实情景的虚拟数据，这些数据可以用 t 检验的方法进一步分析。有时对原始数据的检查也是一件有意思的事情，那可以让你了解被试的真实反应，但缺乏数学上的精确性。统计分析能够指出各组之间的差异是统计显著的。数据支持了绿色的酒让人讨厌的观点，但实验心理学家在陈述这一结论时需要有更谨慎的表述方式。唯一可以确认的是两组之间在机遇水平上（低于 1%）是存在很大差异的。下一步就要进行推论了，研究者可能得出结论，认为原本深红色的酒比加工后变色的绿色的酒更受喜爱，这种偏爱可能受文化的影响。这样，研究的推论就完成了。

这种范式在实验心理学中使用普遍，只要样本够大，代表性足够好，就能认为可能影响实验结果的被试变量（假设包括大量品酒高手）在两组中是随机、等质分布的。

模型 2：匹配组被试设计

有人可能有理由相信某种被试变量（如品酒经验）会对结果产生重要影响，所以被试应当在这个变量上进行匹配。其假设是有品酒经验的人的辨别能力较高，能够进行更精确的判断。这种设计称做**匹配组被试设计**（matched subjects design）或**配对设计**（matched pair design）（模式 2），如下表所示。

同样，被试被标识为 S_1、S_2、S_3……S_{16}，但加入另一种下标 a、b……h，比如 S_{1a} 和 S_{9a} 表示两个被试在某个基础上匹配。本例将被试在品酒经验上进行匹配，每对被试具有大致相同的经验水平。另外，本例中还可以只选择品酒专家，排除牛饮者。可以想象，有些人的喝酒经验很多，但并不一定是葡萄酒的鉴赏高手。其他实验还可能要求匹配其他变量，如性别、年龄、智力、奔跑能力或训练情况等。如果匹配变量与因变量有关系，这种设计就很有效。对匹配变量的测量也很重要，假设要匹配品酒经验或智力水平，如果使用了一个无效测验进行测量，这种设计的基本假设就无法实现。

表 5—2　　模型 2：匹配组被试设计

	实验组（人工调色后的酒）数据		控制组（原色酒）数据
S_{1a}	2	S_{9a}	4
S_{2b}	3	S_{10b}	5
S_{3c}	2	S_{11c}	5
S_{4d}	1	S_{12d}	5
S_{5e}	1	S_{13e}	4
S_{6f}	2	S_{14f}	4
S_{7g}	3	S_{15g}	5
S_{8h}	1	S_{16h}	4

品酒分组：　自变量——酒的颜色
因变量——质量评价
匹配变量——品酒经验

附录 A 描述了配对设计使用的一些计算程序，这种检验有时也称做**相关样本 *t* 检验**（correlated *t*-test）或**相依样本 *t* 检验**（dependent *t*-test），因为这两组在某些重要维度上是相互关联的。对虚拟数据的分析也说明了这两组差异的显著性。

模型 3：被试内设计

被试内设计也称做“重复测量设计”，被试接受两个或两个以上的实验条件。在示例中，每个被试都品尝了两种酒，进行了自身的控制，如被试 S_1 品尝了实验条件（调色后）和控制条件（未调色）的酒。

由于每个被试都进行了自身控制，每个被试能够进行两种条件下的观测，因此这一设计能够收集到更多的数据。同时也能够缩减被试量，通过一个小样本进行研究。不过，在重复测量设计中，被试接受处理的顺序是一个额外的影响因素。比如在品酒实验中，被试对第二次品尝可能会比第一次更有偏好，但这可能并非品尝的结果，而是因为在一般情况下，被试在做第二次品尝时心情会稍微放松或愉快些。所以，就需要进行**顺序平衡**（balanced sequence）。这一研究中顺序的抵消平衡可以让一半的被试先品尝调色后的酒，而另一半被试先品尝原酒。表 5—3 对这一设计进行了描述。

从中可知，被试 $S_1 \sim S_8$ 都首先品尝并评价控制条件的酒，之后是实验条件的酒。而被试 $S_9 \sim S_{16}$ 则采取相反的品酒顺序。通过统计程序，能够分析出这种序列效应。

模型 4：因子设计

先前的几种模型都考虑了单一自变量对因变量的影响。而心理学家通常对几个自变量在因变量上的效应感兴趣。第 4 章中已经指出因子设计在实验心理学中的重要地位，这里再次介绍因子设计中被试变量的使用。在示例中，我们假设研究者选用了酒的颜色和葡萄种类作为自变量，酒的颜色有原色和人工处理后的颜色，酒的类型有三种（选用不同的葡萄酿制，这里假设是黑比诺酒、馨芳葡萄酒和雪当尼酒）。这一设计可以用一个 2×3 矩阵表示（参见表 5—4）。

表 5—3　　模型 3：被试内设计

被试	控制组（原色酒）数据	被试	实验组（人工调色后的酒）数据
S_1		S_1	
S_2		S_2	
S_3		S_3	
S_4		S_4	
S_5		S_5	
S_6		S_6	
S_7		S_7	
S_8		S_8	

被试	实验组（人工调色后的酒）数据	被试	控制组（原色酒）数据
S_9		S_9	
S_{10}		S_{10}	
S_{11}		S_{11}	
S_{12}		S_{12}	
S_{13}		S_{13}	
S_{14}		S_{14}	
S_{15}		S_{15}	
S_{16}		S_{16}	

品酒分组：自变量——酒的颜色

因变量——质量评价

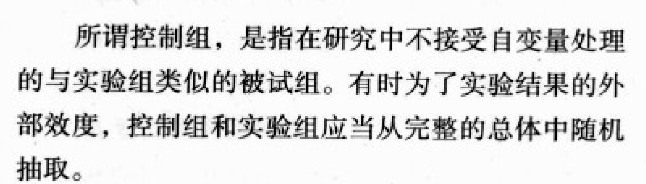
表 5—4　　因子设计：被试内

		因素 1（颜色）	
		水平 1（原色）	水平 2（加工色）
因素 2（酒的类型）	水平 1（黑比诺酒）	S_1、S_2、S_3、S_4	S_1、S_2、S_3、S_4
	水平 2（馨芳葡萄酒）	S_1、S_2、S_3、S_4	S_1、S_2、S_3、S_4
	水平 3（雪当尼酒）	S_1、S_2、S_3、S_4	S_1、S_2、S_3、S_4

同重复测量设计一样，这一设计中的每个被试接受所有条件的处理。这一设计将被试控制技术进行了整合。而其他被试控制技术也是可以使用在因子设计中的，下列模式就描述了独立被试设计在因子实验中的应用（参见表 5—5）。

表 5—5　　因子设计：被试间

		因素1（颜色）	
		水平1（原色）	水平2（加工色）
因素2（酒的类型）	水平1（黑比诺酒）	S_1、S_2、S_3、S_4	S_{13}、S_{14}、S_{15}、S_{16}
	水平2（馨芳葡萄酒）	S_5、S_6、S_7、S_8	S_{17}、S_{18}、S_{19}、S_{20}
	水平3（雪当尼酒）	S_9、S_{10}、S_{11}、S_{12}	S_{21}、S_{22}、S_{23}、S_{24}



5.5　处理组和控制组的使用

所谓控制组，是指在研究中不接受自变量处理的与实验组类似的被试组。有时为了实验结果的外部效度，控制组和实验组应当从完整的总体中随机抽取。

本书已经简要地介绍了几种额外变量的控制方法，让我们回到保罗对演讲恐惧的治疗的研究。研究者希望对两种疗法的效果进行比较，除此之外，研究者还需要控制一些可能随着治疗进度而变化的额外变量。如果某些额外变量随着治疗的进行而发生变化，则因变量上的提高可能归于该变量的变化而非治疗本身的效果。当被试接受某种疗法时，有两个明显的额外变量。首先，已有研究表明有些问题行为的个体在不接受任何特殊的处理的情况下，其症状能够随时间而缓解。这种现象称做**自然康复**（spontaneous remission），也就是说，在没有任何针对该症状的处理的情况下，症状可能自发地消失。在治疗的实验中，研究者要明确被试的进步是由于治疗本身还是由于这种自然康复引起的。

其次，有些被试在认为自己正在接受某种治疗时，其症状可能会有显著的改善，虽然他们事实上没有接受任何治疗处理。这就是**安慰剂效应**（placebo effect）。这一名词来源于拉丁语，原意是“取悦某人”的意思，首先由医生发现。医生提供给病人一些无药理作用的惰性物质（如糖水）来让病人安心，而非出于治理的目的，但医生们却发现某些病人的病情因此好转。心身疾病患者较容易产生安慰剂效应，但也会出现在一些身体的病症上。心理问题的治疗中也存在安慰剂效应，所以治疗组的进步也可能是由于这种效应引起的。

保罗使用控制组的方法来分离自然康复、安慰剂效应和治疗本身的作用。表 5—6 中列出了四个实验组和各组能够分离出的影响因素。研究者通过这一方式很好地分离出不同的变量对被试进步的作用。例如，将各治疗组的提高分数减去安慰剂组的提高分数，就得到了各疗法在剔除安慰剂效应后的作用。

表 5—6　保罗实验的四个实验组及其对治疗效果的解释

	影响治疗效果的变量			
实验组	疗法	安慰剂	自然康复	提高百分比
1. 行为疗法	是	是	是	100
2. 认知疗法	是	是	是	60
3. 安慰剂	不是	是	是	73
4. 无处理	不是	不是	是	32

研究者还可以通过添加处理组的方法以排除额外变量的效应。假设在阿希的印象形成实验中，阿希仅仅使用一个处理组，即先阅读积极的形容词，后阅读消极形容词。这时被试的评价应当是积极的，支持了印象形成中首因效应的假设。但如果只使用了这单独一个处理组，可能会出现以下质疑：首先是阿希的消极形容词可能不够消极，所以被试的积极评价可能并不是形容词呈现顺序的影响；其次是人们可能都倾向于给别人积极的评价，而不在乎他们所看到的积极或消极的信息——即人们可能倾向于寻找他人的优点并接受他人。如果这一假设成立，阿希实验使用不同类别的形容词或呈现顺序都将观察到积极的评价。所以阿希通过实验设计排除了这些可能的质疑，他的一个处理组接受积极 - 消极形容词的呈现顺序，而另一个处理组则接受消极 - 积极形容词的呈现顺序。第一个处理组对虚拟对象的

评价是积极的，而第二个处理组的评价是消极的。这样，先前的质疑和解释就不能成立了，而阿希的两组实验控制也从另一方面说明了这些质疑的不成立。

额外变量控制的原则难以进行简单的归纳。但任何研究者首先要明确有待解决的问题及有待验证的假设；在进行实验设计时，排除任何可能影响结果解释的额外变量。研究者可以保持自变量外的所有实验条件一致，还可以使用更多的处理组或控制组。初学者了解特定研究领域中的额外变量的最好方法就是阅读该领域的实验研究；通过阅读，初学者能够明确该领域中使用的实验设计以及需要控制的额外变量。以下列举了一些存在控制问题的实验研究，初学者能够从中发现过去的研究中出现的一些问题，并有助于分析实验设计和变量的控制。

控制问题：睡眠学习

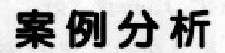
案例分析

本研究旨在确定是否存在睡眠学习。学习的材料是一组德语单词和配对的英语，被试是10名不会德语的大学生。被试睡在实验室的一张舒适的床上，实验室静音且有空调设备。被试在半夜开始就寝，大约在凌晨1:30左右，主试进入房间，询问被试是否睡着了。如果没有回应，主试打开一个播放德语单词和英文释义的录音机。录音材料共60个不同的单词，会持续播放到凌晨4:30。如果期间被试醒过来了，录音暂停播放，直到被试再次入睡后开始重新播放。早上对被试学习的效果进行检测，对被试播放这60个德语单词，让被试报告他们所认为的这些单词对应的英文意义。正确识别的德语单词数目作为研究的因变量测量。实验结果显示，正确识别单词的平均数目是9个（总数60个），而正确识别单词的最高数目是20个。研究者通过解释这些结果，支持了睡眠学习可能存在的假设。

这一实验有几个重要的含义。从理论意义来看，它认为大脑在个体睡眠时仍然是活跃的，可以对外界的感觉信息进行加工；从实践意义来看，睡眠学习是一种轻松的、不需要意志努力的学习方式。对于大学生来说，这绝对是久旱逢甘雨，可以不需要熬夜突击来准备考试，只要简单地打开录音机然后睡觉就行了。

尽管这一研究结论很振奋人心，也很值得期待，但这一研究面临两个主要的质疑。其一是研究没有使用控制组，控制组不接受学习材料而直接参加回忆测试。由于两种语言具有某种程度的相似性，被试虽然自称不会德语，但仍有可能对某些词的含义进行猜测，如德语中的“Mann”的意思就是英语中“Man”。另外，许多德语词汇经常在英语电影（尤其是老电影中）使用。因此睡眠学习的显著效应可能只是被试能够对某些词汇进行猜测的结果，所以需要设置一个控制组考察这一可能原因。

另一种质疑在于研究者对于睡眠的操作性定义。这里的睡眠只是简单地定义为被试在凌晨1:30 ~ 4:30区间内的行为，仅排除了被试自我报告为醒过来的情形。事实上睡眠有浅睡到深度睡眠的很多水平，而在浅睡阶段被试对外部刺激还有部分的意识，但在深度睡眠开始后，对外部刺激的觉察能力就很小了。这一实验中无法得知学习材料是在睡眠的哪个阶段呈现，可能学习只是在浅睡阶段进行，而不是在真正的睡眠水平上进行。



案例分析

西蒙和埃蒙斯（Simon 和 Emmons，1956）设计了一个实验，避免了以上所提的几个问题。实验的学习材料包括 96 种常识问题及其答案，以问答的方式呈现，如：战前格兰特在什么商店里工作？回答是：战前格兰特在一家五金店工作。实验被试有两组，实验组在睡眠时提供了问题答案，而控制组仅是参加了知识小测，而没有事先提供答案。开始实验时，给两组被试提供问题并允许猜测，剔除回答正确的被试；随后在睡眠时呈现给实验组问题和答案，这一阶段，使用脑电图记录被试的脑电波。由于不同睡眠阶段大脑活动是不同的，脑电图记录让研究者能够确定被试的睡眠水平。在每次呈现给被试答案时，被试的脑电波都被记录下来，这样，研究者就知道在提供每个答案时被试的睡眠水平。

早上，实验组就夜间呈现的学习材料进行测验。测验的每个问题包含 5 个选项，被试猜测哪一个答案是正确的。使用多选题的原因是它或许是睡眠学习最好的测量方法，被试在这种测试中只需要再认而不需要再现问题的答案。控制组也参加了同一测试。

获得测试分数后，研究者按照提供答案时被试的睡眠水平将实验组被试进行分类，一共使用了 8 种睡眠水平的分类（在表 5—7 中将之简化为 3 种睡眠水平），报告各睡眠水平下回答的正确率。

研究结果表明，当实验组被试清醒时出现了明显的学习，浅睡水平则有中等程度的学习，但在被试处于真正睡眠阶段时，没有观测到学习的发生。23% 的比率表示了控制组被试猜测的可能结果，而睡眠组与没有学习的控制组的成绩没有差异。

通过这一实验设计，可以看到研究者为了明确判断睡眠学习是否存在而设定了几种严格的控制程序。首先，为保证在睡眠时学习的信息是被试事先未知的，研究者对所有参与者进行预测，排除了已经知道答案的那些被试。其次，由于使用了多选的测试方式，被试的答案有可能是通过猜测得到，所以研究者使用了控制组以了解被试使用猜测时的作答成绩。第三，研究者确定了睡眠的不同阶段以及提供给不同睡眠阶段的答案，这样就能够区分出被试清醒、浅睡或睡眠时呈现的学习材料。通过这些控制方法，实验结果显示，在真正的睡眠水平下没有发生学习活动（参见表 5—7）。

表 5—7　三种睡眠水平下的回答正确率（%）

实验组	睡眠水平		
	清醒	浅睡	睡眠
实验组	92	65	23
控制组	24	23	23

控制问题：社会剥夺和社会强化

案例分析

已有研究不断证实，当动物被剥夺食物之后，食丸作为强化物的作用显著增加。实验希望了解对儿童进行社会剥夺和社会强化时，是否也会出现相同的结果。被试是 6 岁的小学生。

社会强化的作用通过一种弹珠游戏进行测量。游戏使用一种有 2 个孔的箱子，被试每次往其中的一个孔投入一个弹珠。在开始的 4 分钟里，主试只是观看儿童玩游戏；随后 10 分

钟，当儿童将弹珠投入前4分钟里比较少用的那个孔里时，主试给予口头强化，比如说“好”、“真棒”等。实验的因变量是从前4分钟到后10分钟两个区间儿童投入研究者预期的孔中的弹珠增加数。

为确定社会剥夺的作用，被试被随机分派给三种处理。社会剥夺处理时，被试在玩游戏之前要在房间里独处20分钟；非社会剥夺处理时，被试立即开始游戏；社会满足处理时，被试在玩游戏前与主试有20分钟的交谈，并且有画图和剪纸的活动。

实验结果显示，社会剥夺组的被试在往特定孔中投弹珠的增量明显高于另外两组。非社会剥夺组的增量也要高于社会满足处理组。结果支持了事先的假设，儿童社会剥夺或满足对社会强化的影响类似于食物或饮用水剥夺的作用。

这一研究具有广泛的理论意义，它可以解释为社会性动机与基本的生理动机（如饥饿等）可能适用相同的法则。实验的作者们还设计了良好的实验情境来验证他们的假设，实验操作与实验指标的测量都很清晰明确。但在本实验发表之后，质疑产生了，有批评认为该实验的研究结果可能是由于额外变量的控制不当而产生的。

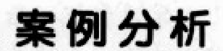
一种质疑认为该研究者在操纵社会剥夺变量时，还引进了一种一般性的感觉剥夺变量。例如，在剥夺处理中，儿童不仅和其他人隔离独处，而且没有玩具可以玩；而在社会满足处理时，儿童不仅和主试交流，而且进行了绘画和剪纸活动。这样，研究者在操纵社会剥夺时便引入了一个额外变量（一般性的感觉剥夺），而实验的结果可能正是由于这一额外变量引起的。斯蒂文森（Stevenson）和奥多姆（Odom）于1962年使用了三组被试检验这一可能的解释。在玩弹珠游戏前，一组儿童独处15分钟并可以玩玩具，第二组独处15分钟但没有玩具，第三组儿童则立即开始弹珠游戏。实验结果显示，两个隔离组的任务成绩并没有区别，但都比非隔离组的成绩高（往特定孔投入更多的弹珠）。由于两个隔离组都属于社会剥夺组，只是其中一组同时剥夺了玩具；而两隔离组的成绩并没有差别。这说明了更高的任务成绩是由于社会剥夺引起的，这也支持了该实验原有的解释，引入的额外变量的影响是很小的。

另有一些研究者也提出对该实验的质疑。他们认为，对于6岁的儿童来说，和一个陌生的成年人在陌生的环境中相处会唤起焦虑。剥夺处理组的儿童焦虑程度应当最大，因为他们要在一个陌生的房间里独处20分钟；其次是非剥夺处理组，这些儿童被直接带到游戏情境中；而社会满足组的儿童的焦虑程度应当最低，因为在与主试20分钟的友好交流后，被试应当会对主试的在场有一定的适应。有研究表明，高焦虑水平会提高某些学习任务（尤其是简单任务）的成绩，因此，该实验的结果可能是由于三种处理唤起了不同的焦虑水平导致的，而非某种特定的社会动机引起的。

为了验证这一假设，沃尔特和帕克（Walter 和 Parke，1964）使用了一个2×2的因子设计。他们使用了两种隔离水平（独处10分钟或直接开始游戏）和两个焦虑唤醒水平。低焦虑水平条件下，主试以友善的态度对待被试，而在高焦虑水平下，主试以冷淡无理的态度对待被试。实验结果显示，（1）两种隔离条件下的行为没有显著差异，这与实验的社会剥夺解释不一致；（2）高焦虑处理下被试的行为显著高于低焦虑处理组，这支持了焦虑唤起的解释；（3）隔离水平与焦虑唤醒水平的交互作用不显著。由此可知，焦虑唤起水平这一额外变量很可能决定了原实验的结果，而社会剥夺的解释可能是不正确的。

控制问题：知觉防御

知觉防御（perceptual defense）对人类行为有着深远意义。这个实验旨在验证知觉防御的存在，并假设个体会无意识地对某个单词是否会引发焦虑进行判断，并阻止或延迟对这类单词的识别。



案例分析

长期以来一直存在这样的观点：人类存在某些自我保护机制，能够防御那些引发焦虑的刺激。知觉防御就是这样一种机制。该实验使用速示器呈现中性词和禁忌词，以验证知觉防御假设。研究者假设禁忌词会引发焦虑，而当被试在无意识水平上进行识别时，知觉防御机制就会延迟被试对禁忌词的识别。基于这一假设，对禁忌词的识别比中性词需要更长的曝光时间。

被试为16名大学生，男女各半。每个被试均单独参加实验，实验时有男女各一位主试在场。11个中性词（如苹果、商业等）和7个禁忌词（如妓女、婊子等）以预定的顺序呈现，使用渐增阈限法测量被试的单词识别阈限。速示器通过快门设计每个单词的曝光速度（0.01 秒），单词的曝光时间逐渐增加，直至被试能够正确识别并口头报告为止。

结果发现，中性词识别的平均阈限为 0.053 秒，而禁忌词识别的平均阈限为 0.098 秒，两个平均数差异统计显著。由于被试识别禁忌词需要更长的曝光时间（更高的阈限值），研究者得出结论并支持了知觉防御假设。

其他研究者很快对这一设计提出批评。豪斯和所罗门（Howes 和 Solomon，1950）提出了两个方法学的质疑。其一，该实验结果可能是因为被试不愿意报告禁忌词，直至他们能够完全辨认该词为止；尤其是面对一个异性主试时，被试可能特别不愿意将这些词口头报告出来。另一个方法学质疑是，中性词比禁忌词在出版物上出现的频率更高，因此被试对中性词的更容易识别可能是由于他们平时更常看到这些词的结果。豪斯和所罗门的后续实验验证了这种词频假设。他们选用了一个词频表，大约有 30 000 个出版物上的词汇；他们从中选取了10个不同频率的词（都是非禁忌词），然后采用与先前案例类似的程序确定单词的识别阈限。他们发现词频与单词识别阈限存在很高的负相关（大约为－0.79），也就是说，单词在出版物上出现的频率越高，单词识别阈限就越低。这个实验说明，词频可能是先前实验中结果的可能解释。那如果控制了词频之后禁忌词的识别阈限是否还会更高呢？

波斯特曼（Postman）、布朗森（Bronson）和格罗珀（Gropper）于 1952 年通过直接测试的方法确定禁忌词和中性词的词频，并将之进行匹配。他们使用了与先前案例实验类似的程序进行研究。其结果没有支持知觉防御假设；事实上，禁忌词的识别阈限要显著低于中性词，这可能是由于过低估计禁忌词的词频所导致的。这一领域的研究还在不断进行中，但似乎支持知觉防御现象的早期的实验结果都是由于词频这一混淆因素导致的。

控制问题：一次尝试学习

一次尝试学习（one-trial learning）认为学习是一次性完成的。以下这个关于一次尝试学习的实验研究有利于理解一种特殊的控制问题。

案例分析

当儿童学习字母表时，通常由老师呈现给学生并读出字母 A、B、C 等，直至儿童学会印刷体字母和发音之间的联系。这种联结是逐渐建立的还是以一种“全或无”的方式形成的？

不同学习理论的支持者对此存在争议。后一种理论流派会认为，如果在几次学习尝试之后，儿童仍不能读出给定的字母，说明字母和发音还没有建立联结；前一种理论流派则会认为，联结已经在某种程度上建立起来了，只是这种联结的强度还不足以让儿童做出正确的反应。

该实验通过巧妙的设计来分析判断哪种理论是正确的。被试的任务是学习 8 对无意义音节，每对无意义音节呈现在一张独立的卡片上。在学习过所有 8 对材料后，让被试根据配对的前一个音节回忆出另一个无意义音节。如，被试学习了 POZ-LER，在回忆测试时，被试需要通过第一个无意义音节 POZ 回忆出第二个音节。

实验使用了两组被试。对第一个实验组，主试替换了每次测试中没有学会的无意义音节配对。例如，被试在学习了 8 个音节配对后接受了一个学习测试，所有测试中没有被回忆出的材料都被放弃，而补充上新的音节配对；然后再次呈现 8 张学习卡片（已学会的和新的配对音节）并进行回忆测试；再次替换未学会的材料，再次进行学习和测试，直至被试能够在一次尝试中学会所有的 8 张卡片为止。另一个实验组的学习和测试方法与第一个实验组相同，但是在所有的尝试中都使用同一份学习材料，直至被试能够在一次尝试中回忆所有的材料为止。

研究者对这两组被试学会所有 8 张卡片所使用的尝试次数进行比较，两组被试进行完整回忆所使用的尝试的平均数完全一致（8.1 次）。由于第一个实验组中的无意义音节配对是在一次尝试中学会的（没学会的都被放弃了），而两组被试尝试的平均数没有差异，说明学习（即联结）是以“全或无”的方式完成的。换个角度来说，研究者认为可以通过第二个实验组来解释联结通过不断复述而逐渐形成的假设；如果复述在学习中有重要作用，第二个实验组对 8 张卡片的学习应当要快于第一个实验组（没有复述）。因为两组尝试次数没有差异，所以可以推论复述是不必要的学习方式。

如同大家所预期的，这一实验也颇让人兴奋，尤其是那些联想学习的支持者（虽然本实验认为联想学习的渐增观点是错误的）。本实验发表不久，质疑的研究就出现了。一个主要的质疑观点是，在第一个处理组中，可能比较难的学习材料都被放弃了，以至于最后的学习材料只包含那些较简单的无意义配对音节；而第二个处理组没有替代学习材料，所以最后的学习材料中同时包含了简单和困难的音节配对。正是由于第二组被试所学的材料要比第一组的最终材料要更困难，所以实验没有发现第二组的学习更快。

1962 年安德伍德（Underwood）、雷胡拉（Rehula）和凯珀尔（Keppel）的实验验证了这种项目选择的假设。他们重复了先前的实验程序，另外还增加了一个控制组。控制组的学习材料为材料替代组（即先前实验中的第一实验组）被试的最后一次尝试的材料，这组被试的每次尝试都使用这一份材料。结果显示，控制组的学习速度比使用所有其他材料（所有材料中随机抽样）的被试的学习都更快。因此在先前实验中第一个实验组所学习的是更简单的材料，这可以用来解释该实验的结果。

5.6　结束语



本书中的案例分析不仅仅用来示范实验原则的应用，还提供机会让你应用所学的知识，让你更精于实验设计并避免问题的出现。在阅读本章中列举的各种控制问题时，请思考其中的实验设计、初始的研究问题、修正后的研究问题、自变量和因变量的确定，以及解释实验设计如何就控制问题进行修正。

第 6 章　被试变量控制

被试变量研究的基本问题是……任何观测到的行为差异都可能是由它们的混淆变量引起的。

——坎特威茨（Kantowitz）、罗迪格（Roediger）和埃尔姆斯（Elmes）

6.1　处理组被试的同质



心理学实验关注特定动物的行为。由于心理学研究需要使用被试，所以对被试特征所引起的额外变量的控制就尤为重要。目前已经发展出许多适用于不同领域的被试变量控制方法，本章将讨论这些控制方法的基本特征。但要注意的是，不少人格和社会心理学家会对个体差异感兴趣。

在研究中，通常将一个处理组的行为表现与另一处理组进行比较。这些处理组中可能存在某些会影响实验结果的被试特征。所以要求所有处理组在这些被试特征上尽可能相同；这样，实验的结果才能够归于自变量的变化，而非由于各组被试某些个人特征（如 IQ）的不同所引起。

现场研究比处理组设计更容易受到被试差异的影响。所有使用自然组被试或自愿参加某种实验处理的被试的研究都要考虑一个重要的问题：这些不同处理下的被试个体差异是怎样的？例如，一个制造业大企业为低级员工举办了一次领导培训课程，这一课程采取自愿参与的方式，并使用自愿参与者的个人时间（晚上）进行；10 年以后，研究者对这一课程的效果进行评价。结果发现，参加该课程的人在公司里的发展比那些未参加课程的人更快。在使用培训课程的效果对这一现象进行解释时，就存在一个问题：当时这个培训课程可能仅仅吸引了那些很希望能够在公司里有发展的人参加。这里的处理组可能是由高动机个体组成，而控制组则可能是由低动机个体组成。因此个体在公司内的发展可能并不是该课程的作用，而仅仅是由于参与者动机差异的影响。

如果研究者使用有关特征基本一致的被试，并将之分派到不同的处理组，这一问题就可以避免。这里有四种常用的方法。第一种是**随机选择**（random selection），随机选择意味着研究希望推论到的总体中的每个个体都有相同的概率参与到该研究中。随机选择只是理想的状况（在某些研究方法中非常重要，如调查研究），但在实验研究中通常难以真正实现。设想一下，如果你接到一个电话，让你在特定时间去参加所在大学里的一个一小时左右的实验，你会有怎样的反应？所以研究者通常进行方便取样（心理系学生就是一个很重要的被试源）或专门取样（经验被试、中风患者、已婚夫妇等）。一旦确定了所使用的被试，就可以使用第二种方法，即随机分派被试到各处理组。随机分派能让研究者在一定程度上相信不同处理组的被试在相关被试特征上基本一致。这种设计可以与先前介绍的实验模型 1 的特征相结合。第三种方法是基于某种相关变量匹配被试（实验模型 2），研究者先就被试在特定任务或测验上的表现进行评分，再将被试分派给不同处理组，以保证不同处理组在这些分数上同质。第四种方法是被试内控制（实验模型 3），每个被试

参加所有的实验处理，这就保证了各处理组在被试变量上的一致。这种设计还需要对同一被试在不同处理中的顺序进行平衡，但顺序效应仍不可避免。后面将分别介绍随机分派、被试匹配和被试内控制的方法。

6.2 随机分派

随机分派是最常用的被试分派方法。**随机**（random）意味着一组被试中的任何个体都有相同的被选择的机会（或概率）。如果实验包含两个处理组，主试可以使用随机数字表或抛硬币的方法分派被试。通过这种随机化程序，研究者可以在相当程度上确信，处理 A 和处理 B 下的被试基本上同质。

这种随机分派程序比较容易掌握，也与随机分派的定义一致，但它也存在一些严重的问题。使用这种程序，研究者最后获得的两组被试可能在数量上不相等。例如，分派结果可能是处理组 A 有 15 名被试，而处理组 B 只有 5 名被试。这不是研究者期望得到的分派结果，显然处理组 B 的 5 名被试所得到的平均数要比大样本下的平均数更不稳定。另一个要考虑的因素是，使用相同的被试数目（n）可以简化实验结果的统计分析。理想的分派程序既要保证一定的分派随机性，同时要保证各处理组的被试数目相等。这种常用的程序虽然并不符合严格意义上的随机分派，但或许更能够保证分派的精确性和无偏性。

在琳达、库珀和琼斯的实验中，被试通过写文章可以获得 50 美分或 2.5 美元的报酬，被试在抵达实验室后被分派给其中一种处理。这种类型的实验通常要延续几周甚至几个月，所以有必要采取一些无偏的被试选择程序；因为较早参加实验的被试与较迟参加的被试可能具有不同的特征。例如，后参加实验的被试可能会从先前参加实验的被试那里了解到有关实验的一些信息，又或者动机较高的被试可能会更早参加实验等。这时要考虑使用**区组随机化**（block randomization）。这种设计使用了一些区组，每一区组都包含了所有的实验处理。例如，在两组实验中，主试可以通过抛硬币的方式分派第一位到实验室报到的被试。假设图朝上则将被试分派给处理 A，字朝上则将他分派给处理 B；而下一位报到的被试则分派给另一种处理。这样，如果第一位被试是“图朝上”，则被分派给处理 A，第二位被试就将被分派给处理 B，这就是一个区组。第三位被试报到时，再次通过抛硬币决定处理组的归属，第四位被试则分派到另一种处理。对整个实验过程重复使用这一被试分派程序，两种处理组在实验的时间区间内等同分布，各区组内都采用了无偏分派的方法。

如果实验多于两个处理，可以用抽纸条的方式进行区组随机控制。把代表各种处理的纸条放在一个容器里，当被试报到时从中抽取一张。例如，假设实验有六种处理，把六张纸条放到一个箱子里，每张纸条上都有一个字母，分别表示六种不同的处理。每个被试报到时，从中抽取一张，被试就被分派给纸条上所标识的处理。这一纸条不再放回箱子，第二个被试则从剩余的五张纸条中抽取。当所有六张纸条都被抽取后，一个区组完成。所有纸条被重新放回箱子，重新开始纸条抽取及被试分派程序。

这种区组随机化程序特别适用于被试会在不同时间段报到的情况。但在另一种研究中，研究者可以同时获得所有被试。例如保罗的实验比较了行为疗法与认知疗法对演讲恐惧症治疗的效果，研究者需要将有演讲恐惧的 67 名学生分派到四种实验条件下。在另一个有关就业指导培训的实验中，指导中心分派 60 人参与该课程；中心主任将其中的一半被试分派给处理组，而另一半被试被分派给控制组。进行这种无偏的随机分派有多种可行的方法，我们只介绍其中的几种。

假设研究者拿到这 60 名被试的名单，希望分派给四种处理组各 15 名被试。抽纸条法可以实现这种分派。研究者将标志为 A、B、C、D 的纸条各 15 张放到一个箱子里，被试的名单以字母顺序排列；但研究者依次念到名字时，被点到的被试从中抽取一张纸条，并根据纸条上的标识被分派给相应的实

验条件，所有抽出的纸条不再放回箱子。这一程序还可以进一步简化，将被试名单以随机顺序排列（例如，把各被试的名字分别印在一张 3×5 英寸的卡片上，然后像洗牌那样打乱卡片的顺序），然后抽取第一名被试分派到处理 A，第二名被试到处理 B，依此类推。

有些实验中被试可能也是同时报到，但研究者不能在实验开始之前拿到被试的名单。假设在一个大房间里有 60 名被试，研究者需要将他们分派给四种实验处理，最简单的方法就是让被试从房间的一头开始报数。将报"一"的所有被试分派给一个处理组，报"二"的所有被试分派给另一个处理组，依此类推。如果被试的就座没有某些特定的模式，这种分派方法也算是一种无偏的选择。通常被试在一个大房间集合后，主试将不同处理条件的测试卡分发给他们。例如约翰逊（Johnson）和斯里普（Scileppi）研究貌似合理的和令人难以置信的信息的传达对个体态度改变的影响。所有传达的信息都以文字的形式呈现在测试卡上，被试在了解这些信息后完成一个态度量表。20 名被试分成两组同时参加实验，实验材料在分发之前先被随机打乱，这样研究者就将两种处理随机分派给所有被试。

类似这些程序能够使更加复杂的分派问题得以轻松解决。例如，在一个小组问题解决的实验中，被试每 3 人一个小组合作解决一种特殊的人际关系问题。自变量是指导语的类型，我们将之称做处理组 A 和处理组 B。这样，研究者不仅要将每个小组随机分派给其中一种处理（A 或 B），还需要随机分派被试给这些 3 人小组。我们让这个分派问题更复杂一些吧，假设每个 3 人小组中要有一个成员被随机指定为小组的领导者。共有 60 名被试在一个大房间报到，研究者需要将这些人分为每种处理下各 10 个 3 人小组，同时随机指定小组组长。这个分派问题可以通过先前介绍的程序轻松解决。首先对每个被试进行编号（每个被试拿到一个号码），主试手上有一套 3×5 英寸的卡片，每个卡片上分别标有 1 ~ 60 的数字。然后主试开始洗牌，并从最上方抽取 3 张卡片，其中与第一张卡片上号码相符的被试成为第一个 3 人小组的组长，而另外两个抽到的相应被试成为该小组的组员。依此类推直至产生所有 20 个小组为止。然后将前 10 个产生的小组分派给处理 A，而后 10 个小组分派给处理 B。这样，一个复杂的分派就以一种合理快捷而且无偏的方式实现了。

6.3　被试匹配 

另一种保证各处理组被试特征同质的方式是匹配法。在匹配法中，被试需要接受某种测试或任务，而这种测试或任务被假设为与真实实验中使用的任务高度相关。被试基于前测的结果被分派给不同的实验处理，这样就保证了各处理组基于前测分数的基本同质。因此研究者也能够保证所有处理组在这个与真实实验表现高度相关的被试变量上的同质。

在讨论匹配法之前，首先要注意进行匹配时的潜在假设。研究者假定他们知道哪些被试变量与被试实验任务上的表现高度相关；另外他们还假定可以通过测验获得所有被试在该特征上的分数。这些假设潜在的风险是：第一个假设可能是不正确的；第二个假设也可能得不到满足，因为对该特征的测量可能是不好的，或者不能通过该测量得到被试的真实分数。

前测任务通常有两种类型。一种任务或测试与实验任务差别较大，但被假定为与实验任务高度相关。例如，研究者可能会假定智力与学习任务高度相关，所以会就智力因素匹配被试。这种前测可能是一个 IQ 测验，它与真实的实验任务有较大差别。另一种前测任务或测验则可能与实验任务相仿（甚至完全一致）。

研究者选定了某种匹配变量并对被试进行测试之后，就可通过一些方法对被试进行匹配。一种方法是配对法。研究者选取了匹配变量上得分相同的两名被试，将其中一名分派给处理 A，而另一名分派给处理 B。假设这是一个两类处理的学习实验，且假定 IQ 与学习任务的成绩高相关。研究者首先获得一个大样本被试的 IQ 分数，然后选取两个 IQ 得分为 135 分的被试，分别分派给处理 A 和处理 B。接着研究者选取两个 IQ 得分为 130 分的被试并分派

给两类处理。依此类推，直至各处理组的样本量达到研究者的期望为止。另一个对精神分裂症患者的药物治疗实验中，研究者希望就被试的年龄因素进行匹配。他可以选取两个 50 岁的患者，其中一位接受药物治疗，另一位接受安慰剂处理；接着他可以选择两个 46 岁的患者并分派给两类处理，依此类推。在两个案例中，两个处理组的被试都进行了完美的匹配，如果实验处理多于两种，仍然可以使用这种匹配的方法。例如，研究者可以选择三名 IQ 得分为 135 分的被试并将他们分别分派给三种实验处理。

这一匹配方法的困难之处在于，要找到分数完全匹配的被试是很困难的，例如，研究者也许只能找到一名 IQ 得分为 135 分的被试。这时研究者可能会使用近似得分进行匹配（如 130 ~ 140 分之间），但即使这样的匹配也很难实现。

事实上，精确匹配被试通常是难以实现的。我们假设一个类似兰伯特（Lambert，1952）和所罗门的大鼠实验，研究者要将 8 只大鼠分派给两种实验处理（A 和 B），并希望就它们在最后一次实验中离开起始箱所耗时间进行匹配。8 只大鼠在起始箱待的时间（秒）依次是 20.5、17.2、10.7、8.0、7.2、6.5、4.3 和 3.2。显然进行精确匹配是不可能的，因为没有任何两个分数是类似的。研究者只能使用一种**等组匹配**（ad lib matching）的程序平衡两组的分数。表 6—1 所呈现的分数区组可能是对这 8 只大鼠的最佳匹配方式。

表 6—1　　等组匹配示例

处理A	处理B
20.5	17.2
8.0	10.7
7.2	6.5
3.2	4.3
$\overline{X}$=9.72	$\overline{X}$=9.67

两组得分的平均数基本一致，且两组都包含了高分、中等分数和低分的大鼠。当然，这种方法也存在自身的问题，因为由此产生的数据难以用恰当的统计方法进行分析。

另一种匹配方法是**随机区组技术**（random blocks technique）。假设在药物治疗实验中需要将 80 位精神分裂症患者分派给四个处理组，同时希望匹配各组被试的年龄，因为有些证据表明年长的患者对治疗的反应更差一些。研究者使用随机区组技术，按年龄排列 80 位患者，将年龄最大的四位随机分派给四种处理；接着分派剩余 76 位患者中年龄最大的四位，依此类推，直至分派完所有 80 位患者。这样，研究者就获得了年龄基本相同的四个区组，并将各区组内的被试分派给特定的一种处理。

如果研究者知道哪些被试变量与实验任务高度相关，并能够获得这些变量的测量，就能够通过匹配的方法有效地排除这些被试特征带来的偏差。

案例分析

梅丽莎·史密斯（Melissa Smith）、伊丽莎白·弗朗茨（Elizabeth Franz）、苏珊·乔伊（Susan Joy）和柯斯蒂·怀特黑德（Kirsty Whitehead）2005 对盲人与视力正常个体的大小知觉感兴趣。招募了盲人被试后，研究者基于性别、教育水平和年龄（2 岁以内）为每个盲人被试匹配了一个视力正常的被试。在实验中，所有被试都被蒙上双眼，要求用双手去抓握并感受十个物体（如奶瓶、汽水瓶、面包块、鸡蛋盒等，所有物体以随机的顺序呈现给被试），每种尝试时间为 6 秒。被试尝试完所有物体之后，主试随机报出某个物体的名称，要求被试使用双手描述该物体的大小（胳膊肘弯曲，伸出双手和手掌），然后使用毫米卷尺测量被试所描述的大小。结果发现，视力正常的被试高估了物体的大小；而相比而言，盲人被试对物体大小的判断更精确。研究者得出结论认为，这是由于盲人使用了更多的精细动作表征，而视力正常个体更依赖精确的视觉记忆表征。

6.4 被试内控制

被试内控制通过让所有被试接受所有实验处理来保证各处理组在被试变量上的同质性，这种设计称做重复测量设计或被试内设计。后一种称谓比较符合该设计的特征，处理间的比较是在同一个被试内进行的。有一个声音频率与绝对响度阈限的关系研究使用了这种实验设计。在实验中，每个被试检测了十种频率水平（如 25、50 等），响度曲线由每种频率水平上四个被试的绝对响度阈限的平均值构成。被试内设计通常用在这一类实验研究中——被试需要对不同情境下的不同刺激作多次判断，这些不同的刺激可以被看做研究的自变量。

被试内设计能够确保各处理组的被试特征一致，但各种处理或刺激呈现的顺序会带来新的问题。实验结果的变化有可能是这种呈现顺序的影响，因为被试对某个特定刺激进行的判断往往会受到先前判断的影响。例如，被试对某个物体轻重的判断一定程度上受到对前一物体轻重判断的影响。因此，在辨别任务的实验中，被试会随着判断次数的增加而变得更为精确与熟练。这通常被称做**练习效应**（practice effect）。另一种对立的现象是**疲劳效应**（fatigue effect），被试可能会随着实验的进程感到疲劳或厌倦，因而造成效率的下降。这些效应可以用几种方法进行控制。

一种方法是**拉丁方设计**（Latin-square design）。假设要让 15 个被试对判断三种不同的刺激（A、B 和 C），拉丁方设计控制刺激的呈现顺序，使每种刺激对每组被试都只呈现一次，且不同组被试接受刺激的顺序都不相同。正如表 6—2 所示。

表 6—2　拉丁方设计

	刺激呈现顺序		
	位置 1	位置 2	位置 3
1.　5 名被试	A	B	C
2.　5 名被试	B	C	A
3.　5 名被试	C	A	B

这里，五名被试以 A、B、C 的顺序接受刺激，五名被试以 B、C、A 的顺序接受刺激，另五名被试则以 C、A、B 的顺序接受刺激。可以看到 A 刺激在第一、第二和第三位置都只出现一次，B 刺激和 C 刺激也一样。最后比较所有 15 名被试对三种刺激反应的平均数。这些平均数排除了顺序效应，因为各个刺激在所有位置上出现的次数是相同的。研究者如果对顺序效应感兴趣，还可以通过被试对不同顺序位置上的同一刺激的反应进行检验。

另一种控制顺序效应的方法是随机区组设计，这一方法在本章已经介绍过。假设实验要求每个被试对七种不同的刺激各做十次判断。这七种刺激可以视为 1 个区组，区组内通过某种随机化程序控制不同刺激的呈现顺序。被试对呈现的第 1 个区组刺激反应之后，七种刺激再次随机组成第 2 个区组，依此类推，直至被试对每个刺激都进行了 10 次判断（10 个区组）为止。所有被试都重复这一实验程序。

被试内设计的优点是，由于所有处理使用了相同的被试，研究者就能保证各组处理在被试变量上的一致性；另外，实验需要的被试量也大大减少了。但在许多实验中不能应用被试内设计，这和采用的实验处理有关。例如，一个实验要对隔离喂养和集体喂养的大鼠进行比较，显然同一被试不可能同时参加这两种处理。再如一个研究要比较高智商和低智商的被试，或法语教学的两种方法，显然各处理应当使用独立的被试组。所以有些实验使用独立组设计会更容易。被试内设计通常适用于被试需要对不同情境下的几种刺激进行多次反应的情况，而这些刺激可以认为是实验的自变量。另外，一些对顺序效应感兴趣的研究也需要使用被试内设计。例如，研究者如果想了解被试从高奖赏情境转换到低奖赏情境后会发生哪些变化，就需要让被试同时参加这两种条件的处理。



案例分析

众所周知，给幼儿阅读是件有益的事，还能够激发幼儿的阅读兴趣并可能帮助其提高阅读技能。玛丽·埃文斯和琼·圣奥宾（Mary Ann Evans 和 Jean Saint-Aubin，2005）对儿童在成人给他们阅读材料（故事书）时的注视特点感兴趣。他们是否在看文字、图案、还是两者都有？书中文字的位置是否会影响儿童的注视区域？他们通过一个眼动实验来探究以上这些问题。实验由一个成人给儿童阅读不同的故事，被试为 5 个四到五岁的儿童，通过前测确定他们都不认识故事书中的文字。故事在文中出现的位置不同，一本书的文字是在该页的上方和下方，中间位置为插图；另一本书的文字总在插图的左边；另一本书的文字出现在插图上的泡泡图案中，这三本书的插图均为彩色。还有两本书的文字总在插图的左边，但插图简单而且没有色彩。给这 5 个孩子阅读五本书的顺序使用了拉丁方设计。对实验结果进行 2（文字或图案）×5（书的类型）重复测量方差分析（ANOVA）分析。结果显示，不论文字出现在什么位置，儿童都很少花时间在文字上；儿童主要注视书中的插图，尤其是色彩更为丰富的插图（与只有简单线条画的书籍相比）。

6.5 被试缺失

我们已经讨论了保证各处理组被试变量基本一致的一些方法。在实验中与被试有关的另一个问题是各处理组被试的损耗或缺失。被试的缺失通常发生在需要被试参加多个阶段实验的情况，例如参加了第一阶段实验的被试并不一定能出现在下一阶段的实验中。以下是一个虚拟的研究，其中就存在被试缺失的问题。

案例分析

有个就业指导中心提供一次为期四周的培训课程，指导年轻人求职的技巧和程序。这个课程需要每天上课，主题涉及去哪里找工作、找哪些工作、怎样写简历以及填写员工选拔的心理测验等内容。一共有 60 名年轻人报名参加，该中心的主任从中随机选择了 30 人参加该课程，另外 30 人作为控制组，以检验该课程的效果。控制组的这些人只被告知班级座位有限，让他们自己求职。研究的因变量是课程结束后一个月内两组被试找到工作的百分比。在完成课程的 16 人中，有 12 位（75%）在一个月内就业；同期内控制组有 15 人（50%）就业。该中心主任通过这一结果来支持其培训项目的效果。

对本研究进行评论之前请大家注意一点，这类研究总是难以做得足够完美。但这类研究问题又是相当重要的。许多培训项目都需要对项目的效果进行评价，比如各种治疗、咨询、矫正等。这些在现实生活中都是非常流行的项目，但很少有研究对其效果进行检验。

本案例的焦点在于，最终的数据统计仅仅来源于最终完成课程的 16 名被试（参加时的人数为 30 人）。实际问题是：哪些被试缺失了？他们可能比较缺乏找工作的动机；或者是能力比较低，理解不了课程的内容；又或是情绪比较不稳定，不能坚持每天到中心上课。就业前景较差的这些个体因此被排除出处理组。另一方面，在控制组是难以进行这种筛除的，所有的就业前景较差的个体仍然待在控制组。该中心主任或许是在拿这组就业前景不错的处理组（不好的已经被排除）与控制组进行比较。如

果这种解释成立，该课程的效果就很值得怀疑了。

实际上，这种解释是否成立不得而知；但它确实是该研究结果的可能解释。一旦实验中出现被试缺失，特别是当某种处理比其余处理的被试缺失更严重时，就应当探究缺失的那些被试的特征。例如，可以观察缺失的那些个体的就业率；如果仅有 25%，就能够支持那种就业前景差的被试缺失的假设；如果有 50% 左右，那么这个课程还是相当有效的。

被试缺失的一个主要危险就是特殊子群的出现，例如，低智商的被试可能会退出某个处理组，但仍保留在其他处理组中。这时各处理组就不应被认为是在被试变量上保持一致了，最终的研究结果可能是由于被试差异而非处理差异造成的。所以获得尽可能多的缺失被试的信息是非常重要的，从获得的信息中能够判断这些缺失被试是否属于某特殊子群。如果信息显示这些退出的被试属于原处理组的一个随机样本，研究者就能够确信他的研究结论没有受到某种特殊的被试差异的影响。但如果信息判断显示这些退出者属于一种特别子群，可能需要排除其他处理组中的这种子群，以保持各组一定程度的可比性。

有时研究者为了避免被试缺失的问题，会让被试只参加一个阶段的实验。如果被试需要参加两个以上阶段的实验，就需要通过某种程序在实验前告知被试这一点，并要求被试在能够参加另一个阶段实验的前提下参加实验。通常一个电话也可能会提醒那些可能遗忘的被试。在某些需要被试在 10 天之后进行再测的实验中（比如某些知觉实验），通常会在被试完成一定阶段的实验后才给予报酬。动物实验通常不存在被试缺失的问题，因为那些被试往往都生活在实验室里，随时都可以进行实验。不过，疾病或死亡也可能会造成被试的缺失，尤其是使用了一些剥夺或压力处理的时候。

案例分析

菲利普·拉什顿（Philippe Rushton）和特鲁迪·邦斯（Trudy Ann Bons）于 2005 年对同卵双胞胎和异卵双胞胎的择偶和择友方式感兴趣。具体来说，他们希望知道人们是否会选择与自己类似的社交伙伴。研究者给双胞胎们发放问卷，要求他们自己填写一份，让其配偶填写一份，让其最要好的一位同性朋友填写一份；问卷内容包括了人口统计学因素、态度和人格因素。这样每对双胞胎都应当收回六份问卷（双胞胎 1、配偶 1 和好友 1；双胞胎 2、配偶 2 和好友 2）。研究者从方法学和统计学上处理研究中出现的被试缺失问题。从方法上，再次邮寄一份问卷给那些没有应答，但其双胞胎有应答的被试，这增加了 25% 的被试参与。从统计方法上，缺失数据使用该项的常模数据代替（常模来自于同性别的所有数据，也就是说，如果某位女性在某个问题上的回答缺失，则使用研究中所有女性在该问题上的反应平均数代替）。研究结果显示，人们倾向于选择与自己类似的配偶和好友。

先前介绍的一些方法以尽可能地避免被试缺失的问题。但只要研究需要较大的时间跨度，被试缺失是必然的，所以收集所有被试（包括那些中途退出者）的信息是非常重要的。理论上这些信息应当包括：（1）实验处理前的信息，例如被试的智力、动机、调节能力等；（2）被试实验过程中的信息，如学习或操作的数据，中途退出者在退出时的发展程度等；（3）处理后的信息，例如本章实例中的就业率等。通过这些信息，研究者更容易对实验结果进行解释。

6.6 结束语

掌握了基本的、高级的设计方法，以及自变量、因变量、额外变量和被试变量的控制技术之后，你现在可以对一些实验研究进行评论，以验证你所学的知识了。

第7章 实验设计的评论

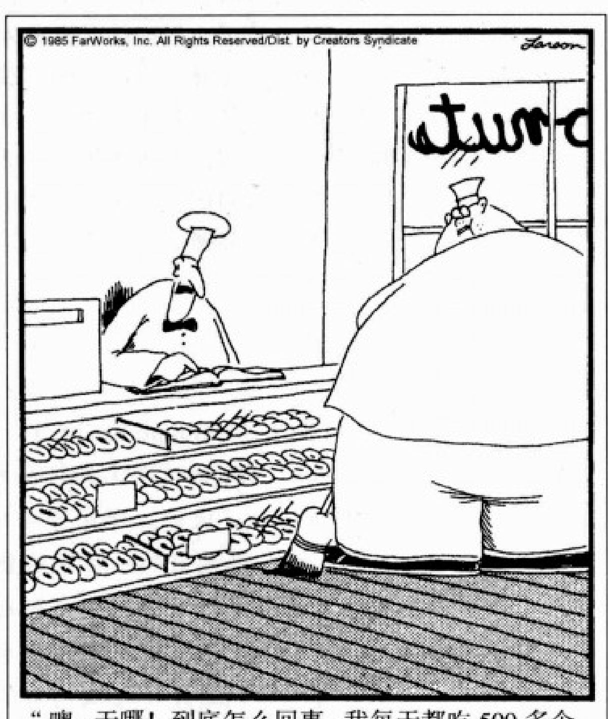
熟能生巧。

——佩里安德（Periander）

心理学研究通常始于对相关研究文献和信息的搜索。研究者首先确定一个需要进一步研究的主题，并形成研究假设；接着研究者会设计一个有效可行的实验。在准备好所有实验材料和被试之后，研究者就能够收集数据、分析数据、进行讨论并最终得到结论。整个研究过程看上去很简单，但每个阶段都潜伏着一些研究者易犯的错误。

在这一部分及以下几部分的内容中，我们将详细分析一种许多资深的研究者或新手都容易陷入的研究陷阱。这种研究陷阱主要是由于研究设计或结果解释时的逻辑不严谨造成的。有些学生把这些问题称为“需要多个备择解释”，因为实验结果很可能有很多原因或解释，而不仅仅是由于某个人所确定的解释（参见下面的漫画）。经验让我们相信，学生对这些错误的练习越多，他们就越能够发现自己研究中的设计或结果解释存在的问题。学生们除了能够阅读并发现本书中所呈现的问题和漏洞外，还能够弥补在自己研究中存在的问题。

本章列举了一些研究概要，每个概要中都存在一个或多个设计问题。这些问题往往发生在研究者如何操作实验，以及如何从研究结果进行推论。我们先前已经介绍了实验设计的基本原则，现在我们的目的是呈现一系列虚拟的研究，并让学生应用所学到的实验设计的知识对这些研究进行评论。这种研究概要是暴露各种研究领域中不同问题的一种快捷手段。评论这些研究不需要太多专业知识，学生可以依照先前介绍的设计原则来发现研究中存在的问题。



“噢，天哪！到底怎么回事。我每天都吃500多个油炸圈饼，为什么还是饿得吱吱叫。”

图7—1 请给一个可能的解释

在评论这些实验概要时，所有学生应当只使用提供的信息，而不需要作任何的推论。例如，假设研究者使用了一个纸笔焦虑测验，在没有提供反面信息的情况下，我们就假定该纸笔测验是有效且可靠的。在本章列举的每个研究概要中都至少存在一个主要的缺陷，学生们应当就这一主要问题进行评论。另外，在评论时必须详细具体；例如，学生不能只是简要地说研究者应当使用一个控制组，而应当具体指出该如何处理控制组。

下面使用一个例子说明如何对研究概要进行评论。



案例分析

有个研究者假设海马（大脑的组成部分）与复杂的思维加工有关，而与简单思维加工无关。他切除了20只随机取样的大鼠的海马结构，让其中的10只学习一个非常简单的迷津，让另外10只学习一个非常复杂的迷津。第一组大鼠通过10次尝试就学会了走迷津（完全没有错误反应），而第二组大鼠则需要至少30次实验才成功。根据这些实验结果，他认为该研究假设得到了证实，切除海马的大鼠在学习复杂任务时比学习简单任务更困难。

这个实验设计符合模型1：被试间设计，所有被试都仅被随机分派给一种处理。

在评论这个设计时，我们很容易发现，任何大鼠学习复杂迷津都可能比学习简单迷津需要更多次的尝试。因此这些实验结果可能和海马结构的切除没有关系——具有海马结构的大鼠可能也会得到同样的结果。换句话说，尽管研究者想要分析两个自变量（任务的复杂性和是否切除海马），但在研究中只操纵了一个自变量。因此评论时应指出，实验应当使用2×2因子设计：一个因素是海马是否切除，另一个因素是迷津的复杂性。这一设计如表7—1所示。

表7—1　　2×2因子设计

	简单迷津	复杂迷津
海马完好	5只大鼠	5只大鼠
海马切除	5只大鼠	5只大鼠

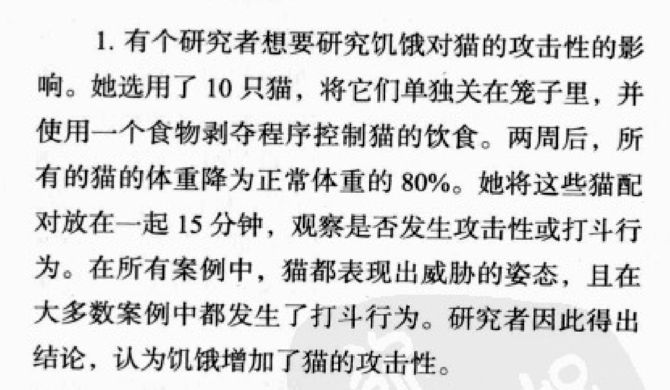
新的设计符合模型4：因子设计，能够评价两个自变量的作用。修改后的设计对研究者假设的验证更为合理。尽管修改后的设计可能比原设计更复杂，但矫正了原有设计的缺陷。库恩曾经用拼图问题来比拟科学研究。他认为拼图“能够检测问题解决的独创性和技能”，而科学研究就如同拼图问题的解决。不妨尝试用你的问题解决（或玩拼图）技能来解决本章中呈现的这些问题设计。

下面一些问题或许有助于你对各研究概要进行评论。

1. 自变量是什么？它至少有两个水平吗，或是使用了实验组和控制组设计？如果没有，说明存在设计的缺陷。任何实验都需要将一种处理与其他处理进行比较。
2. 自变量是被试变量还是可操作变量？
3. 因变量是什么？如何进行测量？
4. 假定自变量有两个水平，是否除自变量外的所有条件保持一致？如果没有，说明实验中存在变量混淆。
5. 实验使用了哪种设计？

另外还可以问些什么问题呢？

7.1 实验概要



1. 有个研究者想要研究饥饿对猫的攻击性的影响。她选用了10只猫，将它们单独关在笼子里，并使用一个食物剥夺程序控制猫的饮食。两周后，所有的猫的体重降为正常体重的80%。她将这些猫配对放在一起15分钟，观察是否发生攻击性或打斗行为。在所有案例中，猫都表现出威胁的姿态，且在大多数案例中都发生了打斗行为。研究者因此得出结论，认为饥饿增加了猫的攻击性。

2. 在食品或饮料公司工作的心理学家通常在产品开发中扮演重要角色。有一个消费心理学家进行了一项实验，测量了被试对两种品牌可乐的喜好程度。该公司发现，它的可乐在某个市场的销售情况明显差于其竞争对手。这是一个让人困惑不解的现象，因为这种可乐在全国范围内的业绩要显著优于其竞争对手的可乐。

研究者认为可能存在某种特殊的地域因素引起

对这种可乐的抵制，并设计实验来验证这种假设。实验采用重复测量设计，每个被试都要品尝两种可乐。一种可乐标志为Q（竞争对手的品牌），另一种标志为M（自己的品牌）。从14～62岁的市民中随机取样，所有被试先品尝Q品牌的可乐，然后再品尝M品牌的可乐，然后再进行喜好评价。出乎研究者意料的是，被试的报告表明他们对M品牌的可乐有着绝对的喜好。研究者推断被试喜欢该公司的可乐，其竞争对手的可乐在该地区的消费必然是广告的作用。于是，他们建议公司开展一次数百万美元的广告活动，以扭转地区的销售形势。

3. 一个研究者想要检验无意义音节学习时集中与分散练习的效果。他使用了三个处理组，将被试随机分派给这三种处理。第一组只练习了一天无意义音节表；第二组连续练习两天，第三组连续练习三天。无意义音节共20个，各组的练习时间均为30分钟/天。研究者使用自由回忆法评估被试的学习效果。结果显示，第一组的回忆平均数为5.2个，第二组为11.0个，而第三组为14.6个，各组平均数在0.01显著性水平上差异显著。研究者因此认为分散学习比集中学习的效果更好。

4. 某个研究者想要探究大学生失败的原因。她选取了一组因表现不佳而被迫退学的“前”学生和一组在大学中成绩优异的学生。她让这两组被试完成一份自尊测试，结果发现，退学组学生在自尊测试上的得分显著低于成绩优异组。她得出结论认为，低自尊是大学失败的原因之一；低自尊的学生很可能会预期失败并表现出失败主义者的行为，并最终导致了失败。

5. 有个研究者设计一个研究，旨在确认高血压患者是否可以通过生物反馈技术控制自己的血压。有一个血压记录装置与每个患者相连，当血压水平上升时，记录装置会发出逐渐增高的声音作为反馈；当患者血压水平下降，则以逐渐下降的声音反馈。患者的任务是尽可能让这个反馈声音保持在安静状态。五个高血压患者各接受了10次生物反馈实验（每次为半小时）。在10个阶段的训练之后，所有五位患者都明显降低了自己的血压水平。研究者因此宣称这种方法是有效的。

6. 这个实验是要验证高动机被试对简单任务的学习比低动机被试更快的假设。进一步的假设认为，在困难任务的学习中将出现相反的情况，低动机的被试将学习更快。研究者对“动机”的操作定义是被试在显性动机量表（Manifest Drive Scale）上的得分。20个高分被试（高动机）和20个低分被试（低动机）需要完成一个困难的学习任务。结果显示低动机组比高动机组的学习速度更快，研究者得出结论，认为该假设成立。

7. 研究者希望验证一个假设：对表现不佳时会受到惩罚的恐惧会不利于动作操作成绩，而不是起促进作用。研究者使用动作稳定性测验作为动作操作成绩的测量，该测验要求被试将一个细针插入孔中，且细针不得碰到孔的边缘（类似九洞测验）。每个被试要将细针插入15个不同的孔中。研究者威胁被试说，在他们表现不佳时会进行电击，以此操纵对惩罚的恐惧。主试会在进行实验前，在每个被试的腿上绑上电击装置，但实际上不管被试表现如何都不会给予任何电击。被试被随机分派到两种条件下：一组被试接受的威胁为50V的电击，是低恐惧条件；另一组被试接受的威胁是100V的电击，是高恐惧条件。与原假设不符的是，高恐惧组被试的表现并不比低恐惧组差，实际上两组的表现基本没有差异。根据这些实验结果，研究者得出结论：惩罚的恐惧对操作成绩的作用微乎其微。

8. 有个研究者选取了20位相信占星术的被试，根据占星术推算各个被试前一天的遭遇，让他们在一个6点量表上评定这一推算的准确程度，6点量表的范围从“极其准确”到“极不准确”。所有20位被试都报告这一占星术推测有一定程度的准确性。因此该研究者得出结论，认为占星术是准确的。

9. 研究者使用2×3因子设计来评估某种实验药物对精神分裂症的治疗效果。实验使用了两类患者：（1）刚进某精神病院就诊的患者，（2）在该院至少住了两年的患者。两类被试都没有该药物治疗的经历。药物治疗的每日剂量有三个水平：3克、6克和9克。每一类患者中选取的被试都被随机分派

给其中一种剂量水平；实验共有六组被试，每组中有20个精神分裂症患者。主试除了监控患者的药物服用外，每周都要对患者是否出现精神分裂症症状进行评定。两个月后，长期住院的各组患者中只有很少数（10%）出现症状的改善，但一半左右的新患者在三个剂量水平的药物治疗中都有改善。研究者得出结论：（1）药物仅对新患者有效，对慢性患者无效；（2）每日3克的药物剂量就能达到最佳治疗效果。

10. 临床心理学中的一个有趣争议是：精神疾病是遗传的还是由环境和经验引起的。有个环境决定论者假设：和患有精神疾病的父母一同居住的儿童，在未来的生活中更容易出现同样的问题。为了检验这一假设，他选取了1 000名成人精神病患者和1 000名正常成人，并考察这些人的父母是否患有精神疾病。只有不足1%的正常成人的父母患有精神疾病，而超过30%的精神疾病患者的父母也患有精神疾病。研究者因此结论，精神疾病并不是遗传的，而是受到儿童期与精神疾病父母的生活经验的影响。

11. 一些研究者认为，如果在每日配给的食物量上增加20%的蔗糖水，大鼠在跑阻力轮时的速度会显著加快。控制组大鼠每日只有固定配给的食物量。实验选用100只来自普利茅斯罗克饲养中心（以提供竞跑大鼠而闻名）的五月花大鼠，随机分为两组（用抛硬币的方式决定）。只有固定食物配给的50只大鼠先跑，处理组的大鼠后跑，结果发现第二组大鼠的奔跑速度更快，这就证实了在高糖的情况下大鼠的精力更充沛。研究者将这一结论推广到小学生中，认为应当让孩子们吃一些高糖的零食。

12. 最近，有个代表新泽西州警察的协会抱怨，由于警察在执勤时使用雷达测速器探测车辆的行驶速度，因此他们患癌症的几率出奇的高。该协会对州政府提出诉讼，宣称那些使用雷达测速器的警察患癌症的几率比那些车辆管理机构中随机抽取的政府职员要高出18%。比较样本与这些警察在教育水平、年龄、性别和服务年限上都进行了匹配；样本容量也很大，警察有238人，车辆管理人员有231人。请问这个案例的理由是否充分？

13. 最近报纸上有篇文章宣称“凉烟糖影响孩子”，还认为买过凉烟糖的孩子比没有买过的孩子更有可能学会吸烟。这个调查指出，在七年级的学生中，买过至少两次凉烟糖的学生要比那些从没买过的学生更可能去尝试真正的香烟。这个清晰的研究结果让很多学校希望能够禁止这种凉烟糖的销售，他们担心这种凉烟糖会引起青少年的吸烟行为。该调查还指出，如果父母亲至少有一方在家中抽烟，孩子就会更倾向去购买凉烟糖，并且会在某些场合尝试吸烟。请问学校的反应是否正确。

14. 有个统计老师想要比较两种初等统计的教学法的优劣。一种方法着重分析统计方法的理论逻辑（称做理论教学法）；另一种则传授学生各种统计检验的方法并解释各种检验应用的情境（称做烹调教学法）。该老师发现，有一所一流的工程学院对所有的初等统计课程都采用了理论教学法，另有一所州立师范学院对所有班级都使用了烹调教学法。在学期末，该老师对这类班级进行了一次标准化统计学测试，结果发现，接受理论教学的班级的成绩远远好于接受烹调教学法的班级。该老师得出结论，理论法是比较好的教学方法，统计老师应当采用这种教学法。

15. 在一项检验氯丙嗪对于精神分裂症的治疗效果的研究中，两个临床研究者从一所精神病院中随机选取了20名急性精神分裂症患者。要求这些被试按照某种维度对几个刺激排序，例如按重量对八个物体排序，有多个类似的排序任务。研究者使用被试内实验设计，首先给所有被试注射生理盐水（安慰剂），然后开始排序任务；在休息几个小时之后再注射氯丙嗪，再次进行原来的排序任务。结果显示，当使用氯丙嗪处理时，被试犯错减少，研究者据此认为，该药物对这类患者的认知功能有较大的促进作用。

16. 有个假设认为感觉剥夺会抑制动物的智力发展。研究者使用了两只大鼠，每只大鼠都刚刚产出8只幼鼠。将其中一只大鼠和它的一窝幼鼠放在一个大笼子里喂养，有足够的空间和事物供其探

索；另一窝幼鼠则与母亲隔离开，且每只幼鼠都被放置在一个很小的笼子里，它们所能看到或听到的只有笼子的四壁与笼子中的喂食器。五个月后，训练长大后的两组大鼠走多分叉的T形迷津，并以食物作为奖励。经过20次尝试，所有非剥夺处理的幼鼠都能毫无错误地跑完迷津；但被剥夺处理的幼鼠仍然会犯不少错误，甚至经常呆滞不动，需要戳它才会继续移动。研究者得出结论，认为感觉剥夺会抑制智力的发展，因此感觉剥夺组的大鼠不具备学会一个简单迷津的智能。

17. 在第二次世界大战期间，有个研究者试图验证惩罚比奖赏对人员训练更有效果的假设。实验的检测任务是判断敌机与友机。她让被试坐在模拟的雷达屏幕前，敌机或友机的轮廓会在屏幕上闪现较短的时间（1秒）；被试按分别标识为“敌”和“友”的两个键对每次闪现的飞机轮廓进行反应。每个被试要接受连续5天共两个小时的实验。第一天的实验中，对被试的每次反应提供正误的反馈。第二天起，被试被随机分派到两种处理，A组的被试每次正确辨认都能获得10美分的奖励，且在错误辨认时没有任何惩罚；B组被试在每次错误辨认时都要接受电击惩罚，但在正确辨认时没有任何奖励。第三天和第四天重复第二天的实验程序。第五天的实验作为测试实验，所有被试的实验程序仍然不变，但不再给予任何奖励、惩罚或其他信息，以每100次刺激的正确辨认次数作为各训练效果的指标。如同预期的那样，在5天的实验中出现了一些被试缺失，A组被试缺失了5%，而B组被试在5天内共缺失了35%。实验结果显示，在第五天的测试中，A组平均正确辨认次数为80，而B组平均正确辨认次数为92，两组平均数差异统计显著。研究者的结论认为，实验假设得到验证，并建议所有的训练项目采用惩罚的方式。

18. 有个小镇的基督教青年会（YMCA）人员想要获得一些证据，以支持他的项目对培养未来的领导者有效。他回溯研究了该项目的成员记录，获得了20年前那些在该项目表现活跃的男孩的名单，同时也获得一些当时没有参加YMCA的男孩名单。他对这两组的当前职业、薪酬等进行比较，发现YCMA组的成就更高。他因此得出结论，这些差异是由于他的项目的影响。

19. 有位心理学家希望开发出一个测验，能够预测律师将来的成功。她从《北美名人录》列举的律师中抽取了一个随机样本，并假定这些律师都是成功律师。她邮寄给这些律师一份包含几百个问题的问卷，对这些问卷进行分析，并描绘成功律师的特征轮廓。然后她再将这份问卷发给一群法律专业学生，并建议那些问卷得分与成功律师特征轮廓差异显著的学生不用再去尝试法律方面的职业了。

20. 有位心理学家提出一个理论：随着小组成员之间相互了解程度的加深，小组的工作效率会逐渐提高；当工作效率达到一个临界点时，会开始轻微地下降。这一下降（“蜜月期结束”）会发生在小组成员停止高度合作行为，而开始争夺权力的时候。为了检验这一理论，这个心理学家选取了互不相识的被试组成小组，让他们完成一系列的任务。研究使用了五个任务，每个任务需要35分钟完成，在两个任务之间每组有5分钟的休息时间。结果显示，小组的工作效率在第五个任务之前一直上升，但在第五个任务上出现明显的下降。据此研究者宣称他的理论得到了支持。

21. 有个临床研究者设计实验比较面谈与客观性测验两种方法对患者病情诊断及确定留院时间的作用。实验在一所大型精神病院进行。其中一组由10个临床心理专业学生分别对6位新病患进行面谈，每次时间为1～2小时。另一组中，有60位病患接受了一系列的标准化心理测验，测验结果由三位有数年经验的临床心理学家进行解释；每位心理学家各解释20位患者的测验结果。面谈或测验组都要求列出每位患者的主要问题，并对该患者的疾病作诊断分类（例如进行型精神分裂症、反应性精神分裂症等）；另外还要预测患者在病情好转之前需要在医院待多长时间。结果显示，面谈组在疾病分类上的正确率为67%，而对住院时间预测的准确率为22%；测验组在疾病分类上的正确率为83%，而对住院时间预测的准确率为65%。研究者因此得出结论，认为面谈法对病情诊断和结果预期都没有什么

价值，不应继续使用。

22. 一个研究者希望验证男性创造力高于女性的假设。她还假设通过控制自我卷入水平，男性的创造力优势会更大。实验使用 2×2 因子设计，一个变量是性别，另一个变量为高或低的自我卷入水平。自我卷入的操纵方式如下：一半被试被告知，该实验是进行一次智力测验，被试的得分将被公布在布告栏中（高自我卷入）；另一半被试则被告知，该实验是为了检验测验的信度，采取无记名的方式（低自我卷入）。创造力测试使用了“特殊用途”测试，每个人就给定的一个物体名称（如锤子），在 5 分钟内尽可能多地写出该物品的特殊用途。两个卷入水平各有 25 名男性和 25 名女性被试，男被试均来自于后备军官训练班，女被试来自于北美女大学生联谊会。实验使用了两个物体：军用罗盘和活动扳手。实验结果显示，男性在低自我卷入下平均列举出 4.1 种特殊用途，在高卷入水平下为 7.6 种；女性在低自我卷入下平均列举出 3.2 种特殊用途，在高卷入水平下为 2.4 种。主效应和交互作用检验均显著，研究者宣称该结果支持了她的假设。

23. 有个研究者需要对三个城市进行快速调查，了解选民所关注的政治议题或问题。这一调查的结果要提供给一个候选人作为竞选活动的依据。该研究者从电话簿中随机选择名单，并尽可能在周一和周二的上午九点至下午五点之间打电话给他们。该结果经整理后提供给候选人，并假定能够有效代表这三个城市的选民的态度。

24. 一位心理学家发现，在乌克兰切尔诺贝利核泄漏事故之后，那些曾经暴露在放射性尘降物中的人们的婚外性行为明显增加。有些受访者报告了灾后的恣意狂欢行为，有些甚至出现奇异的性活动。该心理学家据此推论，放射性尘降物中可能有某种特殊的春药。于是他申请一项研究课题，通过大鼠来验证他的假设。

25. 一位实验心理学家与篮球教练合作，研究运动知识与视觉反馈对篮球运动员和大学生的影响。实验中，被试向镖盘投掷飞镖。无反馈条件下，当飞镖离手时，灯光熄灭，所以被试看不到投掷的结果；反馈条件下的灯光始终亮着。实验后所有被试均要完成一份问卷，测量其对动觉反馈机制的知识。结果显示，无论是否有反馈，运动员的飞镖成绩都要优于普通大学生，但运动员在书面作业上的成绩较差；两组被试在反馈条件下的成绩都要优于无反馈条件，且交互作用不显著。研究者得出结论，认为更多的动觉理论的知识并不会提高动作技能，而对于动作结果的知觉则会提高动作技能。这一结论是否正确？为什么？

7.2　结束语

熟能生巧！对文献进行批判性地评论，有助于进一步提高你的理解和实施实验研究的能力。下一章的内容旨在保证你的实验研究符合一定的伦理规范。

第 8 章　实验研究的伦理问题

诚信至上！

——理查德 · 惠特利（Richard Whately）

纵观美国心理学的历史，研究的伦理问题始终是关注与争论的焦点。心理学实验研究的目的是为了增加我们对人类自身心理特征的科学认识。因此，心理学家通常在实验中使用人类或动物被试。有人认为，有效的心理学法则的建立需要欺骗被试，甚至被试可能受到一些生理伤害。这种观点的解释是，科学知识的探索应当一往无前，而心理学研究所揭示的真理也许会对整个人类有极大的利益。但心理学家仍然要遵守一定的伦理规范，保证被试的心理和生理安全。

许多心理学实验都能够在保证被试的身心安全的同时很好地获得心理特征的科学认识。研究者应当随时注意研究中可能出现的伦理问题，并依据法则，对潜在的利弊进行平衡。

有时研究者获得实验结果的方式存在一定的问题。对人类和动物被试的滥用引起了心理学家和社会公众的极大关注。因此美国心理学会（APA）成立了科学与职业伦理委员会，经过多年的努力，制定出一系列伦理道德准则。本章首先陈述有关心理学家的伦理道德准则，然后通过案例分析加以说明。对这些研究案例，我们可以进行分析或讨论，其中有些明显违背了伦理道德准则，而另外一些则不明显。APA 发表的文件提供了更多伦理问题的实例与解释。

8.1　心理学家的伦理道德和行为规范

这里介绍的是 APA 于 2003 年采纳的行为规范，适用于教学、研究和应用。我们原文呈现了规范中的导言、一般原则和规范中与研究直接相关的一些部分。规范全文可以参见 http://www.apa.org/ethics/code2002.html。使用动物的研究还可以从 APA 的动物研究和伦理委员会获得更多的相关评论，网址是 http://www.apa.org/science/anguide.html 。

导言

心理学家的主要任务是：增加科学、专业的行为知识，促进人们对自身和他人的了解，并运用这些知识来改善个体、组织和社会的状况。心理学家尊重并保护民权、人权，尊重并保护在研究、教学和发表中质询和表达的自由。他们努力帮助公众就人类行为做出合理的判断和选择，因此他们在扮演着许多社会角色，如研究者、教育家、诊断专家、治疗师、监督人、顾问、管理者、社会干预者以及鉴定专家等。本道德规范为心理学家开展他们的专业及学术工作提供了一套普遍的原则和标准。

本道德规范旨在提供能覆盖心理学家所遇到的大多数情况的明确标准。它的目标是，为心理学家的合作对象或群体提供福利和保护，对会员、学生和公众提供关于学科伦理标准的教育。

心理学家的职业操守与伦理标准的形成需要个人的承诺和终身的努力：使自己的行为符合道德规范，鼓励学生、助手、雇员和同事的伦理行为，与他人一起磋商伦理道德规范问题。

一般原则

这个部分由一般原则组成。一般原则与伦理标

准不同，其本质是一种理想的状态，希望能够引导并激励心理学家达到最高的职业道德理想。另外不同的是，一般原则不代表着义务，也不构成实施制裁的基础，以一般原则作为义务或制裁的基础是对其意义和目的的歪曲。

原则A　有利和无害原则

心理学家应努力让工作对象获益，并注意不要造成伤害。在专业活动中，心理学家应设法维护工作对象及其他受到影响的个体的福利与权利，包括研究使用的动物被试。当心理学家的职责和道德义务出现冲突时，他们应设法以一种负责的方式解决这些冲突，从而避免伤害或使伤害最小化。由于心理学家科学且专业的判断或行为可能会影响他人的生活，所以应警惕和预防那些可能导致滥用这种影响力的个人、经济、社会、组织或政治因素。心理学家要意识到，他们自身的身心健康会影响他们帮助工作对象的能力。

原则B　忠诚与责任原则

心理学家应和他们的工作对象建立相互信任的关系。他们应意识到自身对社会及其所在的工作团体负有专业责任和科学责任。心理学家应维护行为的专业标准，明确自身的专业角色和义务，为自己的行为承担相应责任，并设法处理那些可能导致剥削或伤害的利益冲突。为使工作对象得到最大的好处，心理学家可以与其他专业人员或机构进行磋商、借鉴或合作。他们应关注其同事的科学行为和专业行为是否合乎伦理规范。心理学家还应无偿奉献他们的一部分专业时间。

原则C　诚信原则

心理学家应设法提高心理学科学研究、教学及实践中的准确性、诚实性和真实性。在这些活动中，心理学家不应当剽窃、欺骗、参与舞弊、掩饰或故意歪曲事实。心理学家应信守承诺，并尽量避免做出轻率或模糊的承诺。有时使用欺骗可能会在道德意义上使被试利益最大而伤害最小。这时心理学家应认真分析是否应当使用欺骗、欺骗可能造成的后果以及如何弥补使用欺骗可能带来的不信任或其他伤害性结果。

原则D　公正原则

公平与公正是要使所有人都能够受益于心理学的贡献。心理学家应进行理性的推论，警惕自身的潜在的偏好、能力范围以及专业局限性可能导致的操作不当。

原则E　尊重人的权利和尊严

心理学家应当尊重所有人的尊严和价值，尊重个人的隐私权、保密权和自主权。心理学家应当认识到，要采用一定的防护措施来保护那些由于弱点可能会妨碍自主决策的人或团体的权利与福利。心理学家应注意并尊重文化的、个人的和角色的差异，包括年龄、性别、性别认同、人种、种族划分、文化、民族起源、宗教、性取向、缺陷、语言、社会经济地位等。心理学家在面对这类群体中的成员时，应当考虑这些因素。心理学家应当努力消除这些因素对他们的工作所造成的偏差，而且不能有意地参与或忽视存在这类偏见的活动。

8．研究和出版

8.01　机构许可

如果需要机构的审批，心理学家需要提供详细准确的研究计划，并在开展研究前得到相关机构的批准。实际的研究要和所批准的研究相一致。

8.02　研究中的知情同意

a. 如果被试接受了知情同意，心理学家要告知被试以下信息：（1）研究的目的、预期时间和研究程序；（2）在参与研究之后，被试仍有拒绝参加和退出研究的权利；（3）被试拒绝参加或中途退出可能带来的后果；（4）可能会影响被试参与意愿的可预期因素，例如潜在的风险、不舒适感或研究的副作用等；（5）可预期的研究效益；（6）保密的局限性；（7）对参与者的激励；（8）解释研究及被试权利的联系人，他们为准备参加实验的人提供解释。

b. 使用实验处理进行干预研究的心理学家应在研究开始时向被试解释以下信息：（1）处理的实验特性；（2）可能的话，告知控制组能或不能接受的处理；（3）处理组和控制组的分派方式；（4）如果在研究中出现被试不愿继续或希望退出研究的情况时的备用方案；（5）可能的话，要求退出者赔偿或

支付其参与的耗费，偿付资金是否能通过该被试或第三方追索到。

8.03 研究中声音和图像记录的知情同意

在记录被试的声音或图像数据前，心理学家要征求该被试的知情同意。除非：（1）该研究单纯是在公共场合的自然观察，且声像记录的使用方式预计不会带来身份识别或伤害的问题；（2）研究设计使用了欺骗这种控制，而在研究之后的任务释疑时征求被试对声像记录的知情同意。

8.04 来访者或患者、学生、下属作为被试

a. 如果心理学家使用来访者或患者、学生及下属作为研究的被试，应当采取措施避免这些“准”被试因为成绩不良而婉拒或中途退出实验。

b. 如果被试出于课程的要求或为了获取额外的学分而参加研究，这些“准”被试应当有几种等值的活动可供选择。

8.05 研究中知情同意的免除

心理学家只有在以下情形中可以不必征得被试的知情同意。

a. 研究一般不会给被试造成压力或伤害，通常属于下列的研究情形：（1）在教育环境下正常的教育实践、课程或课堂管理方法的研究；（2）匿名问卷、自然观察或档案研究，且对被试反应的使用不会让其有刑事或民事责任的风险，也不损害其经济信誉、就业、个人声誉，被试的信息被严格保密；（3）在组织环境中对工作或组织效率的研究，研究对被试的就业没有影响，被试的信息被严格保密。

b. 联邦或政府法律法规允许的情形。

8.06 诱导被试参与实验

a. 心理学家应避免提供过多的或不正当的金钱或其他诱导方式，以诱导被试参加实验，因为这些诱导刺激可能起到强迫被试参加研究的作用。

b. 如果心理学家以专业服务作为被试参加研究的刺激时，要解释清楚这些服务的内容、风险、义务和局限性等。

8.07 研究中的欺骗

a. 心理学家一般不在研究中采用欺骗被试的手段，除非他们确定该研究具有重要的科学、教育或应用的价值，且无法通过无欺骗的程序进行研究。

b. 如果研究可能导致被试的生理痛苦或严重的精神损伤，心理学家不得欺骗准备参加实验的被试。

c. 心理学家应当向被试解释，研究中的欺骗是设计与实验的一部分。这一解释越早越好，通常在被试刚结束任务时进行，并且不能晚于数据收集完成；允许被试收回自己的数据信息。

8.08 任务释疑

a. 心理学家应及时、恰当地告知被试所参与研究的内容、结果和结论。如果被试存在错误的认识，心理学家应当采取恰当的措施进行修正。

b. 如果由于科学或人文价值的原因必须延迟或禁止提供这种解释信息，心理学家应当采取恰当的措施降低可能的风险。

c. 如果心理学家意识到研究程序对某个被试构成伤害，他们应当采取恰当的措施将伤害降到最低。

8.09 研究中动物的使用和人文关怀

a. 心理学家应以符合联邦、州或地方法律法规及职业标准的方式获得、使用、处置动物被试。

b. 受过研究方法训练和有过照顾实验室动物经验的心理学家负责监督动物实验的所有程序，并保障这些动物的舒适、健康和人道待遇。

c. 心理学家要保证在其监督下的所有动物实验人员都接受了相应的研究方法以及动物照顾、喂养、处置等训练。

d. 心理学家应尽可能减少动物被试的不适、感染、疾病及痛苦。

e. 只有在没有其他可选方案且研究具有重要的科学、教育或应用价值时，心理学家才使用动物被试进行疼痛、压力或剥夺的程序。

f. 心理学家使用动物被试进行手术时，要采取麻醉和适当的方法避免动物受感染，并尽可能减少动物手术中和手术后的痛苦。

g. 如果要终止动物的生命，心理学家应按照合理的程序快速操作，尽可能减低动物承受的痛苦。

8.10 研究结果的报告

a. 心理学家不得伪造数据。

b. 如果在已发表的数据中发现重大的错误，心

理学家需要采取适当的措施及时纠正这些错误。他们可以使用更正、收回、勘误或其他适当的方法进行公布。

8.11　剽窃

即使是偶尔引用，心理学家也不能将他人的著作或数据中的一部分挪为己用。

8.12　版权归属

a. 心理学家对自己实际完成或做出主要贡献的作品需要承担责任，同时享有信誉，包括著作权。

b. 作品的主要作者与其他署名作者应真实反映其对作品的贡献情况，而不受个人的地位的影响。仅作为机构领导，如系主任，不得指定著作权的归属。对作品较少的贡献或对作品出版的贡献可以在脚注或引言部分作适当介绍。

c. 除特殊情况外，学生通常作为在其博士论文基础上发表的论文的第一作者。院系导师应在研究和发表论文阶段与学生沟通论文的署名问题。

8.13　数据的重复发表

心理学家不得把已发表过的数据作为原始数据发表，除非在发表时进行了适当声明。

8.14　分享数据以供验证

a. 在研究成果发表之后，如果有专业人员或机构要求重新分析己方数据和并验证该结论的正确性，只要对方能够保证被试信息的保密性并只用于此目的，且没有相关规定不得泄露该数据，那么心理学家就应当提供该数据。心理学家可以要求该个人或机构负责相关信息提供和索取所需的花费。

b. 从其他心理学家处获得的用以重新分析和验证结论真实性的数据只能用于该用途。如果想用于其他用途，心理学家应当首先获得对方的书面同意。

8.15　审稿人

心理学家审阅其他作者提交的用做发表、出版、授权或研究提案评论的文章时应尊重对方的隐私和著作权。



8.2　使用人类被试的研究

使用人类被试进行心理学实验会产生几个特殊的问题。心理学家既是一名科学家，同时也是社会的一员。在对科学真理的积极探索过程中，心理学家可能会卷入自己的研究，而忽略了某些使用人类被试的伦理问题。这是一种非常严重的错误，最终会给研究者和心理学研究带来不良的影响。

心理学早期的研究没有什么伦理道德准则可以遵循，所以主要依赖于研究者个人的道德规范和社会的约束。事实上，当时的某些研究以现在的标准来看是不被允许的，因而现在招募被试时往往要考虑他们是否自愿参与实验。但有些研究者认为这种被试标准太严格了，会妨碍一些重要数据的收集。另外，也可能会出现新的伦理标准。

这一部分内容介绍使用被试研究时应遵循的原则。**制度审查委员会**（institutional review boards, IRB）在通过研究审核时，需要使用这些指导性原则确定被试（包括人类或动物）是否受到保护。第 9 章中也重申了在进行制度审核时，人类（或动物）的保护问题。

使用人类被试的研究

心理学家通常考虑研究是否对心理科学和人类福祉有贡献，并以此决定是否开展研究；在决定研究之后，研究者要决定如何充分利用研究资源进行研究；然后，研究者开展研究，尊重和保证参与者的尊严与福利，同时要了解并遵循国家与地方法律法规和行业规范对研究中人类被试使用的规定。

a. 做研究计划时，研究者有义务对研究的伦理许可性进行细致评估。通过对科学价值与人道价值的权衡，研究者有义务提出伦理方面的建议，并遵循严格的人类被试权利保护措施。

b. 根据已有标准判断计划研究中的被试属于“高风险被试”还是“低风险被试”，这是研究者基本的伦理道德准则。

c. 研究者有义务保证研究过程中的伦理性。研究者还应对其合作者、助手、学生或雇员对被试处理的伦理性负责，所有这些人也应负类似的责任。

d. 除“低风险”研究外，研究者要和被试在参加研究前达成明确且公正的协议，明确双方的责任与义务。研究者有义务兑现协议中的许诺和承诺。

研究者应告知被试那些可能影响他们参与意愿的研究信息，并解答被试提出的其他研究信息。如果不能在达成知情同意前暴露充分的实验信息，研究者要有其他措施以保证被试的福利和尊严。使用儿童或存在理解和交流缺陷的被试进行研究时，要有特殊的保证措施。

e. 研究方法可能要求使用隐瞒或欺骗技术。在进行这样的研究之前，研究者有特殊的责任：（1）要确定使用这些技术有利于研究的科学、教育或应用价值；（2）要确定没有其他不使用隐瞒或欺骗技术的可行程序；（3）要确认能够尽快为被试提供充分的解释。

f. 研究者尊重每个人在任何时候拒绝参加或退出研究的自由。当研究者对被试有权威或影响时，应特别考虑保护这种自由的责任。这种权威包括（不仅限于）以下情形：参与研究是职业的要求，或被试是研究者的学生、来访者或雇员。

g. 研究者保护被试不受研究程序可能造成的身心不适、伤害和危险。如果研究存在这类风险，研究者应如实以告。研究不应使用可能导致被试严重或持久性伤害的程序，除非不使用这些程序可能让被试有受到更严重伤害的风险，或研究有重大的潜在意义并获得每个被试完全的、自愿的知情同意。应告知被试当潜在的研究压力、伤害、相关问题或疑问出现时，及时与研究者联系的方式。

h. 数据收集完成之后，研究者应提供给被试有关研究的信息，并尽量消除被试可能的错误认识。如果出于科学或人文价值的考虑而需要延迟或禁止提供这些信息，研究者有责任对研究过程进行监控，以确定没有对被试造成伤害性的后果。

i. 如果研究程序对某些被试造成了不良后果，研究者有责任发现并消除或纠正这些后果，包括长期的影响。

j. 除非事先得到同意，研究中所获得的被试信息都应保密。如果他人可能会获取这些信息，这种可能性应当在获得知情同意环节和保密原则一起向被试进行解释。

案例分析

史密斯（Smith）、蒂雷尔（Tyrell）、科伊尔（Coyle）和威廉姆（William）于 1987 年在《英国心理学杂志》（*British Journal of Psychology*）上发表了一个实验，通过实验控制感冒和流感以探究轻微的疾病是否会改变人们的工作效率。实验招募志愿者，让他们到一个感冒研究中心生活 10 天。他们以二或三人一组被安置在隔离的房间里。在三天的隔离期后给这些被试注射含有病毒或安慰剂的滴鼻。48 ~ 72 小时的潜伏期后，由医生对所有被试进行检查，以评估病情的严重程度，客观性指标包括体温、用掉的面巾纸和鼻腔分泌物的分析。

接着，所有被试要完成两个操作任务。一个是侦测任务，被试要侦测不定期出现的刺激并快速反应；另一个任务检验被试的手眼协调能力。

结果显示，流感降低了被试在侦测任务上的成绩，但对手眼协调任务没有影响。总的来说，感冒产生了副作用。

该实验使用的程序得到了当地伦理委员会的认可，并获得所有被试的知情同意。被试排除了孕妇和服用安眠药、镇静剂或抗抑郁药的个体。所有被试还接受了一次医学检查，包括 X 胸透，检查不合格者也被排除。被试没有得到报酬，但得到了食物、住宿、路费和一些零用钱。其他临床实验也使用了相同的控制。

请同学们讨论这个研究的伦理问题。

1. 研究者是否遵守了合理的标准（如 APA）？
2. 被试是否受到强迫？
3. 将收益与风险相比，这个研究是否有价值（注意其他使用的测验）？
4. 是否有其他方法可以收集到这些数据？
5. 被试接受的处理是否符合 APA 原则？
6. 问问其他同学，他们是否愿意参加该实验？
7. 他们能否收集到该实验的数据？为什么？
8. 全班讨论这一研究实例（或原文）。

8.3 案例分析



这一部分我们将呈现几个研究项目，其中有些存在明显的问题，而另外一些则似乎符合先前描述的那些准则。请分析这些项目中的缺陷及合理之处。

案例 1

有位心理学家对那些在第六感测试中得分高的人的人格特征感兴趣。实验首先让被试（接受者）报告对另一个人（发送者）正在观看的第六感卡的印象。第六感卡的顺序是随机的，且发送者完全不知道接受者的存在。有几个接受者得分较高，但他们与其他接受者的人格特征并没有差异。尽管如此，研究者认为他已经获得了一组对第六感信号特别敏感的被试，他使用这些被试进行了四个附加的实验。在前三个实验中，被试的得分显著高于几率水平，而第四个实验中被试的表现与几率水平没有差异。得到这些结果之后，研究者决定只报告前三个实验，而将第四个实验的结果归于发送者或接受者的疲劳。

1. 该研究是否存在程序性的问题？
2. 研究的伦理问题是什么？

案例 2

一位接受过研究设计以及生理心理学训练的实验心理学家在东部一所著名大学工作，有个生产健康谷物食品的大公司与她洽谈。该公司希望这位心理学家设计一个实验，说明谷物纤维在减少常见病（如感冒）和降低缺勤率上的效果。如果她愿意设计该实验并同意在其后的广告活动中使用其名字，公司同意支付大额报酬。这位心理学家同意了，并约定她将拥有最终确定广告的权力。她将不直接参与数据的收集，但相信数据的收集会严格依据她拟定的标准。虽然缺少营养学方面的知识或训练，她相信在实验设计与生理心理学方面的背景足以帮助她设计出一个有效的实验。

1. 这位心理学家应该接受这个合同吗？
2. 为什么？

案例 3

在一个道德规范的研究中，实验心理学家对人们道德信念的强度感兴趣。在一个实验中，研究者告诉被试，有个小孩非常需要一种药物，而这种药物要从一种生长在某个石灰岩洞里的真菌中提取。这种真菌在洞中有很多，而治疗只需要很小一部分就够了。然而洞穴的主人拒绝让任何人使用这些真菌。研究者发现很大比例的被试都报告说，他们可以非法侵入去取得这些真菌。第二部分的实验中，研究者直接让其中一些被试以非法途径去取得真菌。通过对这一程序的分析，研究者认为存在一种更高级的道德规范，该实验结果对公民违法行为的理解有重要的影响。

1. 研究者的计划是否符合伦理规范？
2. 这个研究能否进行修正，从而既能达到心理学家的目的，又能符合伦理道德准则？

案例 4

有个研究人类记忆问题的心理学家开发了一种

高级记忆技术。她决定参加电视智力竞赛节目来检验这种技术。多次成功表现之后，她决定写一本关于该技术的书。由于电视智力竞赛节目的表现使她闻名于世，她同意将照片放在书的封面和广告上，同时描述道："你，也能掌握约瑟芬·布朗的独特记忆技术！"

1. 这个案例中存在哪些伦理问题？

案例 5

在一次律师集会上，有人让一位社会心理学家描述她的一个关于陪审团决策过程的研究结果。在研究中，她访谈了一个著名的谋杀案的陪审团成员。在研究中，陪审团各成员的身份都得到了很好的保护，但她描述了团体的详细特征，例如，陪审团中有7位女性，2位非洲裔成员，一位出生于国外的意大利裔成员，一位建筑师和一个卡车司机。研究者还提到他们的投票和协商模式。招致伦理不当的质疑后，她认为自己并没有提到陪审团成员的姓名，并认为这些陪审团成员已是公众人物，他们的意见早已不是秘密。

1. 这里的伦理问题是什么？
2. 你的看法是什么？

案例 6

有位中西部某所著名大学的博士毕业生要完成他的博士论文，他关注母亲的宗教态度与孩子尿床倾向的关系。他的样本为48位20～28岁的白人母亲，她们都来自某宗教团体，她们的孩子都是健康的。在他马上就能完成研究时，有4位被试退出了实验。由于答辩日期很接近了，他决定从有孩子的朋友中招募被试，同时确保所有的新被试的特征符合设计的要求。

论文的答辩很成功，他被授予博士学位。现在他是中西部一所著名大学心理系的重要人员。

1. 这个毕业生的行为是否符合伦理规范？
2. 该实验是否存在其他补救的方法，而不需要重新开始，或从朋友中招募被试？如果有，该如何进行？

案例 7

"我参加过的最好的研究生讨论是在尊敬的史密斯教授的工业心理学课上。"有位同事告诉另一位教授，"讨论会上每个同学都会拿到一个关于空间对工作效率影响的具体假设，我们做了大量的文献综述并设计了一个实验。我们甚至预期了实验的结果并进行分析和讨论。教授提出问题，而我们完成了所有的工作，这真是一种不错的学习经历。"

几年以后，这两位同事在一次心理学会议上碰面。"你还记得我告诉过你的那个史密斯教授的故事吗？"一个人说。

"哦，是的……"另一个回答，"你说过他那个不错的讨论会。"

"得了吧，那个老骗子，"第一个人鄙视地说，"他窃取了我们的研究思路并进行了实施，而且把结果发表在最新的《工业研究杂志》上了。看，这可是抄袭。"

该论文中的设计与讨论组提交的类似，甚至在研究引言部分直接引用了原文中的内容。

1. 这位教授的行为是否符合伦理规范？
2. 这些学生与这位教授应当如何避免这种不快的情景发生？

案例 8

有位社会心理研究者希望使用一个"基本态度"测验观察态度与学业成就的关系。但被试报到并完成该态度测验时，主要研究者缺席了。由于该测验比较容易操作（实际上可以由被试自己操作），研究者决定让他的秘书代为管理。这位秘书没有心理测量学的正规背景，但了解该测验的管理程序。数据收集之后，心理学家对测验与被试的结果进行了讨论。

1. 该程序是否符合伦理规范？

案例 9

目击证人对刑事罪犯的指认是一个重要的社会与心理问题。有位研究者决定策划一次有目击者的犯罪活动，并让目击者对罪犯进行描绘。实验

在一个快餐店进行，所有的店员都事先进行了细致的彩排。罪犯由一个演员扮演，他进入店铺，拿出一只没有装弹的手枪，要求收银台交出所有的现金。他告诉店员不许报警，并且在离开的时候大声叫道："谁第一个离开门口就干掉谁。"这个罪犯一离开，研究者和助手立即进入店铺，并给在场的顾客一份问卷。问卷内容包括罪犯的外貌特征、是否有武器、说过的话等。然后给每个顾客呈现一些照片，让他们辨别这个罪犯。

问卷完成后对每位顾客都进行了认真的任务释疑，并介绍了这一问题的社会与心理意义。另外还提供了进一步释疑和咨询的机会，但没有被试认为需要接受更多的干预。

1. 从伦理角度评价这一实验。
2. 请设计一个既能评估目击者指认，又没有采取欺骗技术的实验。

案例 10

在一个维生素 A 对半黑暗环境下大鼠迷宫学习能力的影响的研究中，研究者相信，最低剂量会促进大鼠的行为，但在更大的剂量下，大鼠行为会出现下降。研究者选取了四种维生素 A 注射水平，已有研究证实其中最高水平的维生素 A 会使大鼠中毒，但研究者认为对假设的验证需要这个水平的处理。

已有研究还认为高维生素 A 会影响大鼠的迷宫成绩，但这一假设没有在本实验中得到检验。因此，实验结果揭示了一些新的东西，具有重要的科学价值。

除了高维生素 A 注射组外，其余大鼠都得到很好的照料。收集到满足分析需要的数据后，实验终止，大鼠被快速无痛地杀死。

1. 研究者是否遵循了动物被试使用的伦理道德准则？
2. 你对本研究有什么看法？

如上所示，准则难以涵盖心理实验工作中可能出现的所有道德问题，正如法律也难以涵盖人类行为的所有偶然事件。不过，我们所描述的准则是为了提供一个能够广泛应用于的特定领域的普遍原则。这些原则会受到解释的影响，或许不同的研究者会有不同的解释。当设计一个可能涉及伦理问题的研究时，研究者在伦理解释前应先征求其他学者的意见。实验的伦理与非伦理有时存在明显的界限，他人的建议或许能帮助你明确某个实验程序是否可行。

8.4 结束语

保证实验的伦理性能避免违反某些规则。也就是说，要在研究的所有阶段保持诚信，要以应有的尊重对待动物和人类被试。

第 9 章 研究过程

霍夫施塔特法则：任何事花费的时间总比你想象的更长，即使在你思考霍夫施塔特法则时也是如此。

——道格拉斯 · 霍夫施塔特（Douglas R. Hofstadter）

本章讲述心理研究的过程，包括如何获取研究经费、如何开展研究以及如何在会议或刊物上发表研究。在某种意义上，这些是实验研究中最复杂的步骤；但只要实验设计足够细致，所有这些步骤就能水到渠成。图 9—1 显示了研究的主要步骤。但这些步骤仅仅是研究的指南。在真正的实验研究中，研究者经常要回溯、重新开始或重新考虑研究假设，阅读新的文献，从头开始研究，重新设计实验，深入分析，寻求帮助，思考新的观点，考虑研究方案的实践性，并在研究的各阶段保持逻辑性的、创新性的思维方式。首先我们来考虑如何做研究。



图 9—1 研究计划、实施与报告的步骤

9.1 如何研究

研究的步骤

在开始一个实验之前，研究者应对该领域的已有发现有相当的了解。很可能你想做的实验已经有人做过了。大量的心理学文献阅读除了能让你知道他人做过什么研究，更重要的是，你能够了解哪些问题需要进行研究。先前研究中的信息也可能会让你产生新的**研究问题**（problem），即有待分析、解决和解释的问题。实际上，研究问题说明了在当前的专业领域存在可质疑的方面或空白。研究问题可以形成**假设**（hypothesis），或可检验的命题。当确定研究问题并提出假设后，研究者就可以认真细致地设计一个实验，以验证假设并解决研究问题。在这一阶段，研究者要特别注意研究设计与实验控制

（参见先前的章节）。研究者要获得被试的同意，开始实验、分析数据、撰写研究论文并完成整个研究过程。

获取研究思路

第 2 章中描述了许多从现实生活和日常经历中获取研究思路的情形。个人经历是你研究兴趣的第一源泉。下一步是回顾文献，了解该主题已有的研究和发现。在你刚开始第 2 章的学习时，就已经开始进行本课程的文献搜集和阅读了；现在，你应当开始准备你自己的设计了。回顾第 2 章的内容有助于你的设计。无论如何，现在你的文献搜集和阅读技巧应该已经比较成熟了，而你的目标应当从专题学习转向一个研究计划。对于研究计划来说，你首先要深入了解相关文献，并严格评定你的研究思路，然后将你的研究思路转化为研究假设。

现实的研究方案

计划一个可行性实验的一大要素就是提出可操作的方案。假设某个人对人脑研究感兴趣，然而他却没有必需的技术和设备；或内陆城市的某个人对鲸鱼的自然交配行为感兴趣。要将这些研究者的思路付诸实践似乎是不可能的。

许多研究方案要用到许多实验室所没有的昂贵设备。但如果研究有重要意义，就可以去寻求研究基金或机构设备款的资助。这一过程可能会耗费大量时间，并且不能保证一定能够得到研究资助。另一个解决方案是直接前往某个能够实现研究方案的实验室（如医学院）或环境（如海洋）。

另一个操作性问题是被试的获得。内陆城市没有鲸鱼。有些地方，研究者甚至找不到可以用来研究的大鼠、大学生、精神分裂症患者、遗尿者或小学三年级学生。而且，在使用被试进行研究时，还应考虑伦理问题（参见第 8 章）。

还有一个问题在于，所申请的研究能否在限期内完成。以我们的经验来看，这一点正是许多学生和他们的教授最为犯愁的。实验计划的实现往往比预期的时间要长得多，原因有很多，例如制度审查委员会的拖延、设备故障、需要重做实验、被试没有参加实验、自然灾害、合作者不可靠、领域中的新发现、停电、放寒假等。好的研究需要大量的时间和投入，著名的科学家都将他们生命中的大部分时间奉献给了科学。

研究设计

在具备研究思路，并将这偶然的想法转化为合理的研究设计之后，你就可以开始设计你的研究了。你可能首先要将研究中的所有概念或因素进行可操作化（参见第 3 章）。第 3、4、5、6 章的内容有助于你的实验设计，同时你还要考虑伦理道德问题。

评估研究的需要和资源

研究计划过程中的重要一步是评估研究所需的材料、设备或其他资源。许多研究问题既可以使用“低技术水平”的方法，也可以使用“高技术水平”的研究方法解决。例如，你可以使用纸笔方式完成一次问卷调查；也可以将它编程后通过计算机呈现，让被试直接输入反应。选择使用哪种方法取决于你的时间、精力、专业水平和所能获得的资源。有些研究问题需要专业设备（EEG 设备、专业计算机软件、眼动仪等），有些则使用相对容易获得的设备，如电视机、VCR 等或简单的材料，如记录板、笔和纸等。现在，越来越多的研究使用了计算机和计算机软件，可以让被试直接输入反应，记录反应时，并精确控制刺激的呈现。由于计算机较为普及且相对经济可行，因而逐渐成为研究实践的“标准”。不过，对设备要求和所需资源的评估仍然是研究计划过程中不可或缺的步骤。

申请研究经费

你现在或许已经意识到，进行研究需要有资金支持。有时研究花费很少，研究者自己能够承担。有时研究者的任职单位（如大学、公司或研究机构）会资助研究。但很多时候，研究者需要寻求一些外部的研究资助，在本书第二部分介绍的许多实验研究都属于这种情形。（请查看这些论文首页对研究资助的声明）

你可以与研究导师合作，获得所在大学或学院的内部研究资助。通常院系或学院会向其学生与教员提供资助，这些资助通常在 100 ~ 500 美元之间，有些学校可能会多或者少一些。不论如何，你都应当努力尝试获得资助，这不仅能够帮助你完成研究计划，还能够向他人（如研究生院、供职单位）显示你熟知这些资助方式并能够成功获得资助。这些机会通常要求提交一份课题申请报告，这与以下描述的外部研究资助的申请过程类似。

外部研究资助的最常见的形式被称做“基金”。对各位的职业生涯来说，考虑这一问题似乎有点早，但熟知课题基金申请的基本制度和程序是极其必要的。另外，基金申请的过程和内容涵盖了研究设计过程应考虑的所有部分，对学生或任何研究者都会有帮助。

我们来看一下基金申请过程的步骤。

1. **形成严谨的研究思路和设计**。包括本书中一直强调的批判性思维和严格控制的实验设计。

2. **确定与你课题研究计划相符的资助机构**。研究者要与能够联系到基金机构的大学、研究机构、医院或大型企业保持紧密联系。这些单位会定时发布大量基金机构的研究信息，并提示那些可能资助你的研究领域的基金组织。如果没有找到基金机构，研究者可以搜索最近的研究图书馆（如与某大学合作的图书馆），并咨询相应的研究指南，了解基金机构列表及各机构资助的研究类型。另外，研究者还可以通过一些政府的研究基金网站进行咨询，如美国国家科学基金（www.nsf.gov）和美国国家健康研究院的隶属研究项目（www.nih.gov）。

为了符合某基金机构的要求，你可能要对原来的研究计划进行修改。这里我们有一点建议：如果你认为研究是有意义的，最好寻找一个最贴切的基金机构，而不是过分屈从于某些基金机构的要求。换句话说，如果你对帮别人做研究感兴趣的话，那你也不妨按照该机构发布的要求设计实验。

3. **获取并填写基金课题申请表**。研究者搜集到相关基金机构的信息后，可以与学校的基金部门或该基金机构联系，并索取特定的基金申请文件。也可以浏览这些机构的网站，因为许多文件都被放置在网站上了。

最重要的是，必须严格依照课题申请指导文件进行！对学生来说这点可能被重复强调了多次，对于研究者来说也非常重要。一个很好的设计可能会因为没有严格依照申请指南而被直接退回。

4. **提交课题申请，并在需要时进行修改**。完成申请报告之后，提交报告以供审核。有些基金机构接受申请的打印稿；但大多都要求使用电子提交的方式。认真检查你的大学科研基金部门和基金机构的网址以获取更多的细节。

研究基金机构使用不同的标准确定哪些申请报告可以获得资助。总的来说，基金申请极有可能，甚至正逐渐成为研究者研究生涯的主要部分。

课题申请书

标题页。标题页（也称做封面或申请表）的内容包括：项目名称、申请人的单位名称和地址、接受申请的机构名称、项目的起讫日期、总预算要求、同意提交的人事授权签名和论文提交日期。

目录。界定课题申请书的各个部分。

摘要。也称做概要或内容提要，对课题申请的主要内容进行简要概述。

引言。也称做研究背景，介绍申请人的供职机构，并介绍课题申请的结构。

问题提出。清晰准确地描述要解决的问题，介绍项目研究的意义、相关性、迫切性、可推广性和可能的效益。

研究目的。也称做一般目标或解决方案，以定性的方式概述预期的结果。

研究目标。也称做可测量目标、预期结果或假设，对项目研究可能出现的结果以研究目标、研究假设或问题的方式具体描述。

研究方法。也称做研究活动、操作计划、工作计划、研究策略或研究程序等，描述如何实现研究目标或验证假设；通常包括研究设计、总体与抽样、数据收集与研究工具、数据分析以及研究的进度表等内容。

研究的发布。也称做结果推广、转让、应用或复制，说明最终结果或发现与他人分享的方式。同时确定应当提交给基金机构的报告。

设备和设施。也称做研究资源，描述研究需要的设备、设施以及提供的方式。

项目组成员。这部分提供该项目组各成员的信息、工作角色、研究背景、研究资格等。包括主要成员的个人简历；新人的比例、类别以及如何进入项目组。还包括介绍项目的组织管理方式；指定作为研究顾问的人员，描述他们的背景以及在研究中的作用。

预算。这部分具体介绍研究项目的花费，通常分为几部分内容，如劳务费、资料与材料费、差旅费、数据处理费、设备设施费用以及其余间接的或管理费用。

附录。附录内容可以有：正文中过长的相关资料，主要成员简历、同意书、图表及参考书目等。仔细阅读资助方的申请指南，因为课题申报有时不使用附录，以限制申请书的页数。

获取制度审查委员会的同意

在开始收集数据之前，你先要获得所在高校的制度审查委员会对你使用人类或动物被试进行研究的同意。作为一名学生，这些工作可以由合作的教授完成。

制度审查委员会是政府法规要求设立的机构，对所有使用人类或动物被试的研究方案进行审查。制度审查委员会通过实施美国联邦政府法规、州法律和学校规定以保障研究的正常进行。制度审查委员会负责对所有涉及人类或动物使用的研究进行初始审核和定期的检查。这些审查的目的是为了保证人类和动物被试的权利与福利，有些大学对使用人类被试或动物被试的研究有独立的审查委员会。

制度审查委员会负责保护研究中被试的权利，这是通过对研究申请的细致审查而实现的，包括：

- 科学设计；
- 研究方法；
- 研究程序；
- 被试总体；
- 招募程序；
- 知情同意程序。

制度审查委员会试图通过这些审查尽可能权衡科学的理论验证、社会价值和被试承担的风险。

为保证研究者对联邦法规的理解，所有课题的第一负责人和项目组主要成员都必须参加并完成一个关于被试保护的在线课程。获得这一资格的基本方法就是完成美国国家健康研究院提供的一个在线训练模块，你也可以到以下网站完成这一训练：http://cme.cancer.gov/clinicaltrials/learning/humanparticipant-protections.asp。

你需要填写一些表格，提交研究课题申请及知情同意的材料。你可以和学校的制度审查委员会联系，获取各类指南和表格（参见下例）。

人类被试知情同意书

项目名称：（与制度审查委员会申请表上的名称一致）

研究者姓名：________________________________

实验邀请：你必须得到一份参与研究的正式邀请。例如“我们邀请您参加由XXX大学进行的一个研究项目。大学要求您提交一份签名的同意书，然后参加研究。以下内容可以帮助你决定是否参加这项研究。”

研究性质和目的：清晰准确地说明研究的目的。

研究程序的解释：说明所有的研究程序，包括其目的、持续时间、频率、是否使用声像记录、研究的最终结果等；包括让被试对将要做什么以及将要被问到什么有一定的认识；包括可能使被试退出研究的

任何情形（除非被试同意）。

不适及风险：说明研究项目可能造成的任何生理、心理、社会、法律及经济上的风险或损失。如果任务风险很低，可以使用诸如不适、压力、不便之类的表述。可以这么说：参加研究的风险是很小的；参加的风险和在日常生活中经历的没有差别；没有可预见的风险。

受益和补偿：说明研究带来的所有直接收益。包括由于实验处理增进了生理或心理健康、提高了技能等。补偿和受益是区别开来的，包括现金、礼物、路费或学分补偿等。如果被试没有直接受益，也应进行说明。可能的话，说明主动或被动退出或终止实验对其受益的影响。注意所有被试群体接受的补偿应当是相等的，而不应以此胁迫被试参与。也就是说，如果是对被试的时间进行补偿，即使有被试在结束前终止了实验参与，仍然应当按比例提供一定的补偿。

保密性：说明对被试信息的保密措施：实验中的信息将提供给哪些人或机构；提供哪部分信息；披露信息的目的。例如："实验中获得的能够识别您身份的信息都将得到保密。不包含任何识别信息的总结性发现可能会被发表在一个专业杂志上，或在学术会议上被报告。"

拒绝或退出的权利：提供自愿参与原则和随时退出而不受任何处罚的解释。例如："您的参与完全取决于您的愿望，您可以在任何时候退出研究，或者一开始就拒绝参与。您不会受到任何处罚，也不会影响您的其他利益。"

疑问：被试可以询问本研究项目的更多信息。例如"如果您对本研究有疑问，希望了解参与之后的问题，您可以联系研究者________，电话：XXX-XXXX 或是____________科室的研究指导____________，电话：XXX-XXXX。您还可以向人类被试协调办公室咨询被试权利和被试审查程序的问题，电话：XXX-XXXX。"

同意声明：包括以下陈述。

我完全认识到参加本次研究的性质和范围，以及可能存在的风险。我在此声明，同意参加该研究。我确认已经获得本次知情同意书的副本。我现在已满 18 周岁。

________________________　　　　________________________

（被试签名）　　　　（日期）

________________________　　　　________________________

（打印的被试姓名）

________________________　　　　________________________

（研究者签名）　　　　（日期）

________________________　　　　________________________

（研究者或指导者签名）　　　　（日期）

注意：将一份完整的知情同意书副本（不仅仅是同意声明部分）交给研究负责人，而另一份交给被试。签署的知情同意书应至少保存 3 年以备查。

研究课题申请

研究题目________________________ 研究者________________________

________________________________ ________________________________

________________________________ 日 期________________________

________________________________ ________________________________

研究主题

____

____

研究问题

____

____

研究意义

____

____

已有研究

理论____

____

____

设计____

____

____

结果____

____

____

实验范式____

____

____

数据分析____

____

____

实验材料____

____

____

研究进度计划与分工

研究阶段

	材料准备	________________	负责人________
阶段 1	文献搜索	________________	负责人________

	（试点）预实验	______________	负责人________
	实施实验	______________	负责人________
阶段 2	数据分析	______________	负责人________
	论文撰写	______________	负责人________
阶段 3	论文发表	______________	负责人________

研究计划

这些年来我们一直和学生及合作者进行研究，我们发现，使用研究计划表进行研究设计和思考是非常有帮助的。以一个研究课题申请为例，它包括了论文题目、研究者、研究主题、研究问题、研究意义；接着提供一部分空间描述已有的研究、理论、设计、结论；接着描述本实验的范式和数据分析的方法；然后对实验材料和被试进行介绍；最后是一个简洁的研究进度表和项目组各成员的职责分工。这是研究计划中非常重要的部分。为了避免合作者间出现矛盾，就应当明确研究团队中各个成员的具体职责。合作者们接着可以讨论哪些人可以成为作者，文章署名的顺序，哪些人会在脚注中提及，而哪些人将不会在成果中出现。如果有两个以上的成员对项目的贡献一样大，署名顺序有时也可以用抛硬币的方式决定，又或者按字母表顺序决定。有时作者们还可以在研究过程中变换论文发表中的署名顺序。这时应当摒弃过于自我的想法，以团队中各成员的实际贡献来分配该研究的荣誉。对这个表格可以进行扩展，它是相当实用的研究计划工具。

实际上，在进行研究的过程中，你应当有随时进行记录的习惯。有些研究项目可能没有经历头脑风暴阶段，有些可能需要更多被试参与，还有一些分析可能不够完善。无论如何，保持这些记录直至研究完成或取得进展。有些人会发现保留研究日志的重要作用。在日志中，你会随时记下那些阅读文献之后，或其他任何时间与活动中闪现的观点，这有助于形成更好的研究思路。

试点实验

即使研究设计已经考虑了所有的细节，仍然需要进行至少一次的预实验、试点实验，或者是**试点研究**（pilot study），这样才能保证研究者对实验技术与实验程序的熟练掌握，同时也能进行方法上的修正和设备的调试。在试点研究中，使用被试的目的是要用来评估或产生研究的刺激（或变量）。通过试点研究收集被试的反馈信息，可以确保研究使用的工具或呈现刺激的方法能让被试理解并进行反应。当涉及欺骗技术（你要确定被试是否相信了实验的表面解释）或被试要完成多个步骤或研究任务时，试点研究就很重要了。你会发现，对预实验被试的讨论将是进一步完善研究的宝贵资源。

实施实验

实验的实施通常需要一个场所（实验室、礼堂、教室、机房）并招募被试。实施实验会面临很大的压力，你要保证自己的条理性。试点研究能够让你不会丢三落四，且具备实践经验。记住，你本人也是实验中潜在的误差来源。所以你要保证在每个实验阶段中的实验程序、指导语或态度都尽可能一致。当错误发生时，要记录下来；你的结果应当如实反映研究的过程。你或你的研究导师也许会因为某些错误而感到非常不安，但尤其重要的是，你应当记录下这些错误并让他人知道这些过程。不要试图在收集数据时文过饰非。哪怕需要重新再做一部分实验，也好过让所有结果由于某个错误而无法解释或毫无意义。这也正是研究者应有的诚信。另外，在对待被试时，请始终保持尊重的态度，并遵从伦理道德准则和一般的伦理标准。

数据分析

下一步是将数据输入到诸如 SPSS 或 SAS 的统计软件包中（假设数据没有自动记录）。大学的学生用机上一般都有这类程序，也可以通过导师获取。

如果你已经瞄准了某个研究所，那买一个学生版的统计软件也是不错的选择。

接着就是数据分析了。描述统计提供了不同类别的统计图，是对数据的描述。推论统计使研究者能够确定所得到的结果能否支持他们的假设，抑或只是属于随机的变异。推论统计能让你通过对样本的测量，得到无法直接测量的大范围总体的推论。根据学习情况，你可以自己完成数据分析，也可以寻求他人帮助，或者由导师完成部分工作。无论如何，这个阶段是你找到研究问题的答案的时候。对某些学生来说，统计思想和数据分析是枯燥无味的，甚至是可怕的。其实不用担心或烦躁。统计学就如同一些数据在讲故事，它们提供给我们孜孜以求的答案。因而对许多研究者来说，数据分析是最激动人心的过程。附录 A 和附录 B 对统计方法有更多的介绍。

9.2 撰写研究论文

实验研究的最后阶段是撰写研究论文；毕竟，没有在科学界报告的结果是没有多少价值的。研究者形成研究问题，设计并实施实验，然后分析数据，现在他们要考虑的是如何将研究发现呈现给他人。

研究论文讲述了一个故事。最复杂的论文遵循的仍然是一般原则，也许你在高中阶段写一些“五段作文”的时候已经运用过。在文章的第一段（或部分），你要告诉读者文章将讨论的问题，后面的一些段落支持了你的前提假定，而最后一段则进行收尾总结。研究论文当然要比五段作文长得多，但其逻辑是一样的。对很多人来说，论文撰写的开始部分是最困难的；在论文撰写之前，应当有一些准备工作，例如重新回顾并组织研究的资料和笔记等。先写出一个大纲，大纲中的每个标题最终都可以扩展成一个段落或一个部分，这是开始讲述你的研究故事并进行写作的好办法。另一种方法是了解并学习科学论文的写作习惯，你可以借鉴一篇已发表的科学论文来指导你的写作。

几乎所有的心理学杂志都使用 APA《出版手册》（2001）中的格式，认真的研究者会获取一份论文格式指南。APA《出版手册》提供了以下方面的信息与指导：（1）论稿的内容与结构，（2）观点的表达，（3）使用无偏语言，（4）APA 编辑风格，（5）论稿演示及样稿，（6）论稿的接受和出版，（7）APA 的下属杂志和（8）论稿清单。

好的科技论文写作需要练习。第一次尝试就非常成功是比较罕见的，所以我们鼓励学生不断进行修改和重写。即使是有经验的作者也要不断地进行修改和重写。这和你的学校作业不同，你可能会写上一篇学期论文（只有一稿），交给老师，拿到分数，之后你可能再也不会去看它；而科学写作靠的就是修改！下面的一些写作指南会有助于科技论文的准备，不过我们仍强调练习的重要性；在发表前，你最好还能让同事们审阅你的作品。

撰写格式

斯滕伯格的《心理学家的伴侣：学生与研究者科学写作指南》（*The Psychology Companion:A Guide to Scientific Writing for Students and Researchers*）（2003）是关于科学写作的很好的指导。他以清晰易懂的方式描绘了科学写作的要点。他认为，与许多人所想不同的是，好的科学写作实际上是一种创造性的活动。它需要提供信息，说服他人，表达的内容与表达的方式一样重要。斯滕伯格提醒我们，撰写过程是思维过程中的重要成分；写作通常能让我们达到对特定对象的清楚认识。斯滕伯格还提出忠告：要记录你的观点并呈现所有的事实，而非仅呈现那些支持你的观点的事实。

科学写作的另一目的是为了进行清晰的思想交流。APA《出版手册》给出了一些建议。

“为了清晰明确，好的写作应当使用精炼的语言，避免观点表达的模糊，力求书面表达的简洁、流畅，并考虑到论文的读者。成功的作者会让读者乐于开始并继续阅读的过程，并能够通过条理清晰的思维引导读者进行愉快的思想旅行。”

好的科学写作要求文字表达的意义明确，避免使用模棱两可的词汇。在日常交流中，我们可以使

用诸如此类的表达方式："大多数情况下"、"很少"、"我估计"、"使用了一个智力测验"、"动物被剥夺了"等；但在科学写作中，作者应当使用能够操作性定义的词汇，例如，"很少"会让被试困惑到底有多少，"使用了一个智力测验"也没有指出使用了哪个具体的智力测验。

另外，观点的表达要有逻辑。这在介绍研究和讨论结果时尤其重要，在描述科学研究方法时也很重要。大家可以设想一篇研究论文的文献综述部分，这通常是论文的第一个主要部分。文献综述这一部分的目的是向读者介绍本研究对象的已有研究，以表示本研究问题的提出乃是基于有意义的研究背景。研究者不可能详述所有的相关研究，所以必须选取最具代表性的相关研究；读者通过这些高度逻辑的叙述，了解先前研究或理论的发展过程（本书第二部分中有许多优秀的实例，读者可以从中学习不同的作者是如何组织文献综述的）。

论文撰写还要特别注意语法和段落扩展问题，避免使用别扭的词序。论文应保证句法的正确，当然还应保证句子能够清晰地表达研究者的意图。

研究者常犯的一个错误是在一个段落中罗列过多的观点。每个段落都应有一个核心观点，段落中的每个句子都是对这一观点的支持；段落间则依照某个原则进行组织。如果主题是"黑屋子里蝙蝠的声学回馈"，或"儿童的梦境与遗尿行为"，或"蓝色和绿色包装对洗衣粉销售的影响"，则各段落应当围绕相关主题展开。

有效沟通

如果研究者清楚地表达了他的观点，有效的沟通就实现了。这首先要求清晰的思维过程，然后是将想法付诸文字的能力。清晰的思维和写作技能都能通过练习和反馈（教授、同事或自我的反馈）而得到提高。在设计一个好的心理实验时，清晰思维和写作技能也非常重要。我们来看一个著名的经济学家写的一段让人发昏的文字。

"展望未来，彼得先生相信，中短期利率会出现 50 ~ 70 个基本点的下调，而长期利率会保持在一个相对稳定的水平。在股本市场，他预测企业生产力的提高将继续成为一个积极因素，并将使公司获得超出预期的利润。"

我们或许会这样理解，"短期利率可能会下调，而长期利率会保持不变；所以公司可以加大生产并赚到更多的钱。"作为表达混乱的练习，请以一种混乱的方式写一段文字，并让同事或朋友来破解你的原意。

使用无偏语言

APA 致力于科学写作的规范，避免出现对个体或群体的苛刻态度或偏见。研究者应当遵循下列一些 APA 的标准，确保写作时的准确和没有偏见。

- **标准 1：精确地描述**。精确是科学写作的要求，当描述某个体或群体时，应使用准确、清楚和没有偏见的词汇。描述的精确程度依赖于研究问题和该领域的知识现状。
- **标准 2：对称谓敏感**。要尊重他人的喜好，以他人喜欢的方式称呼。要意识到个人喜好会随着时间而改变，而团体内的成员间通常对称谓的喜好也不一致。要确定适用的称谓，你可以直接向被试询问，尤其是在小组内各个被试的意见不一致的情况下。
- **标准 3：参与确认**。你需要在研究中采用某种记录方式来确认被试的参与。不要采用冷冰冰的"被试"这一称呼，你可以适当地使用一些略为温暖的词汇，例如参与者、个体、大学生、儿童或受测者等。
- **性别**。在使用名字、代词和形容词具体描述被试时，要避免出现性别特征或性别角色的模糊。使用代词时的粗心大意可能会出现性别偏差，例如男性代词"他"被用来指代两种性别，或专门使用男性或女性代词定义某种角色（如，这位护士……她……）。英文中的"man"被作为类属名词或职业代称的后缀时（如，policeman）也会引起歧义，这可能错误地暗示为该群体中的所有人都是男性。要明确你所指的是某一性别还是两种性别。用"他或她"和"她或他"来取代"他"的方法也应该

少用，因为不断重复会引起读者的厌倦。像“他 / 她”或英文中的“(s)he”的组合方式则会造成尴尬和分心。“他”和“她”交替使用也是不合理的；这么做似乎还意味着“他”和“她”属于不同类别，这与研究事实不符。对各种代词的使用都会不可避免地影响读者的性别判断。

- **性取向**。性取向与性偏好不同。在标准 2 的解释下，除非对选项进行意义解释，否则应使用“性取向”。对于某些特定群体来说，“女同性恋者”（lesbians）或“男同性恋者”（gay men）会比“同性恋”（homosexual）更适用。
- **种族或民族认同**。种族或民族的称谓偏好经常会发生变化。一个原因纯属个人喜好，这些喜好随着人们命名的不同而变化。另一个原因是时代的变化，称谓可能会过时甚至带有负面的意义。研究者应谨记准确性和敏感性这两个标准。按照标准 2，作者可以直接向被试询问他们喜欢的称谓，尽量避免负面的称谓。例如，有些非裔喜欢被称为“黑人”（Black），而另一些则喜欢被称为“非裔美国人”（African American），一般情况下这两个称谓都是可取的。按照标准 1，对样本的描述应当准确，应尽可能使用更准确的词汇。
- **身心残障**。要遵循不使用“残废”这种表述的原则。应避免将被试等同于其境况的表述（如神经病患者、残疾人），避免使用多余的、含弦外之音的表述，或被视为污蔑的表述（如瘸子）。使用“残疾”或“残障”时，可以参照个体的特性或受限制的类别，如在态度、法律、建筑学上或能力本身的限制（如，阶梯或门槛会限制那些需要无障碍坡道的个体）。用“有困难的”或“特殊的”通常显得较为委婉，但仍要被试同意后才能使用。一般来说，“……的人”、“伴随着……的人”、“具有……的人”之类是中性的且较为合适的表述方式。
- **年龄**。在研究方法部分介绍被试时，需要对年龄进行定义。要提供具体的年龄范围，避免使用开放性的定义，如“18 岁以下”或“65 岁以上”。英文写作时，“boy”和“girl”适用于描述高中年龄段或年龄更小的群体，“young man”、“young woman”、“male adolescent”和“female adolescent”也适用。对于 18 岁以上的被试（大学年龄段或更大），则使用“men”和“women”。“elderly”通常不作为名词使用，而且被有些人认为是贬义的形容词；“older person”的表述更合适。

《出版手册》提供了关于语言偏见的更多信息和实例。在撰写论文之前不妨先参详手册中该部分的内容。

论稿的组成部分

在其他方面，作者可以选择适用的写作格式，但在报告的一般结构上没有什么选择的自由。通用的结构保证了文章各个部分对特定问题的解答。在引言部分回答“研究什么”和“为什么研究”的问题，方法部分解答“谁”和“如何”进行研究的问题，结果部分报告研究的发现，讨论部分则解答“谁关注”的问题。心理学的报告有着严格的结构，如下所示（本书第二部分有更具体的讨论）。

标题页
摘要
引言
方法
结果
讨论
参考文献
附录作者说明

下面一一介绍这些主题。

1. **标题页**。这一页应包括标题、作者姓名和单位以及栏外标题。标题应当是研究主题的简洁陈述，应当说明研究的实际变量或理论问题以及它们之间的关系。例如，_________对_________的影响。标题之后是作者的姓名和任职单位（大学或学院等）。标题页上还应当有栏外标题，栏外标题是个简短的标题，能够让看到该论文的读者意识到所处页码属于同一篇文章。样稿中标识出了栏外标题的

位置。

2. **摘要**。作者在摘要中对论文进行概括。摘要字数不应超过120个单词（或200个汉字）。研究者要在摘要中说明研究的问题、方法和结果。摘要应当能引起别人的兴趣，并且提供一定的信息。记住，这是人们在搜索文献时所看到的部分，因此，一个精心准备的摘要可能是论文中最重要的一段文字了。

3. **引言**。你需要用一两段文字解答以下问题：（1）为什么这个研究主题或研究问题很重要？（2）研究假设与设计和该问题有什么关系？（3）研究的理论意义是什么？（4）它与先前研究的关系如何？（5）研究要检验哪些理论命题？（6）这些命题是如何产生的？似乎一两段文字难以涵盖这么多问题。但是，引言中的前一两段话只是对引言其余部分的铺垫，你可以在后续的段落中具体介绍研究背景、目的和逻辑。引言的长度依不同的出版要求或教授的不同要求而各异。有些杂志发表相当简短的报告，而有些则发表较长的论文。你可以通过期刊、杂志的“投稿须知”了解更多的信息。这部分内容不需要标题（你不需要在论文开始打上“引言”二字）。

4. **方法**。这一部分介绍实验设计，描述应尽可能详细，以便重复实验。这里通常要描述四部分内容：（1）参与者或被试；（2）刺激和材料；（3）研究程序。复杂的研究还可以增加刺激或设计部分的内容。方法部分犹如食谱，所有的原料和步骤都必须清楚地呈现，否则实验就难以重复。

- **参与者或被试**。你需要恰当地说明研究的被试。这对于评价研究结果以及重复研究有重要意义。如果使用人类被试，应当报告人数、招募方式、选择与分派方式及补偿或报酬方式，还应报告性别、年龄、种族或民族等人口统计学方面的信息。如果还有其他重要的人口统计学变量，也应当进行报告。如果使用动物被试，则需要报告所属的种、属的信息和提供者的姓名及场所；还需要报告动物数量、性别、年龄、体重、生理条件等方面的信息。如果对它们的处理或操作与研究非常相关，也应当对它们进行描述。
- **仪器、材料与刺激**。这部分描述实验使用的仪器、材料和刺激。一些诸如家具、秒表之类的标准设备就不需要进行具体介绍了。专业设备则需要描述供应商和型号；定制设备则需要详细描述，甚至需要照片或绘图说明。这里你应当详述所采用的测量方法（如问卷、量表、调查等）。你需要列举一些研究的例题和被试的反应模式；如果使用他人的量表做研究，还应当提供引文，包括量表信度的信息。如果你向研究被试呈现了刺激，可以在这一部分进行完整地描述；如果刺激比较复杂，也可以单独列出一个部分，作为研究的刺激部分。

专栏

被试还是参与者

最新版本的APA出版手册指导研究者把人类被试称做参与者。这个变化承认了参与我们实验的人的自愿性、主动性。然而，有人也指出用参与者代替被试显得很笨拙，而且纯属文字游戏。心理学界已经使用被试这个术语100多年了，并且你可以注意到它在本书中的使用，尤其是当讨论这条新惯例出台之前的研究。我们建议你使用不带偏见的语言，并恰当地描述你在研究中所使用的个体（人类或动物）。这可以为你的读者提供科学准确和可重复的信息。不论你选择把研究中的人类称做参与者还是被试，这最终取决于研究者自己。

■ **设计**。并非所有的论文都需要“设计”部分，许多设计问题可以放在“程序”部分解释。不过，如果你的研究比较复杂，就可以增加一个“设计”部分，对研究变量和实验范式进行完整的描述。

■ **程序**。这一部分要提供关于如何进行研究的具体信息，还要描述自变量的操纵、各处理组信息、给人类被试的指导语等。随机化、反向抵消平衡以及其他控制技术也要在这一部分描述。“研究程序”部分应当让读者知道研究者和被试在研究过程中做了什么。如果内容过多、过繁杂，你可以增设一个“设计”部分（参见上一点）。

5. **结果**。这部分应当报告实验的所有相关数据及对这些数据的分析。文中应呈现或提及所有相关数据，但有些数据最好以图表的形式呈现。描述统计是对数据的概括，应当在结果部分呈现；推论统计报告了研究结果归于机遇的可能性，也应当被明确并适当地呈现。对数据录入或所用统计软件的介绍则不是必需的（参见后面的“数据报告”部分）。

6. **讨论**。这部分的目的是对研究结果的意义进行评价和解释。其中包括：指出研究结果的局限性，指出研究结果与先前发现之间的相似或不同，以及研究结果对特定研究领域的理论或实践意义，并建议将来的研究。总之，你的目的是审查、解释和论证研究的结果，并最终由该结果进行推论。你还可以讨论为什么该问题是重要的，该结果与非研究情境下的复杂现象有何联系，以及该结果与哪些真实现象有关。

7. **参考文献**。论文中所有的引用都应在这一部分标明，每一个列举的参考文献都应在文中被引用。一般格式如下：作者姓名、出版年、文章标题、期刊标题、卷号、页码。

8. **附录**。附录用来呈现那些不利于文章流畅性的详细信息。例如对设备的详细描绘、实验室的平面图、未发表的问卷等。可能在论文发表的时候，编辑出于版面的考虑会让你移走这部分内容（代之以“材料”部分的某些陈述：“问卷可以向第一作者索取”）。

9. **作者注**。在作者注中，你要注明各位作者的院系单位、资金资助的来源，向他人的贡献或帮助致谢，并注明研究的联系人，以便读者获取更多的信息。

9.3　数据报告



实验结果的报告形式可以是文字形式、数据形式或图表形式。本书只介绍了少量的数据分析，这不是因为它们不重要，而是因为本书的重点在于心理学的研究设计。其他书籍会详细介绍统计的使用。

观察数据的量化

几乎在每个心理学实验中，研究者都收集到大量的数据，这些数据可以归为两种类别：描述数据和推论数据。简单来说，这两类数据对应着两种主要的实验研究类型：观察研究和实验研究。

如前所述，观察研究包括自然观察、数据研究和相关研究。自然观察可以是记录一只雄性蓝翅莺在交配季节前后的发声频率。数据研究通常是对一个被试样本进行数据描述，并由此进行推论。相关研究则用于确定两个或多个变量之间的关系，例如大学成绩与高中成绩、血压和收入、创造力和IQ等。总之，观察研究的目的是对自然发生的现象进行精确描述。这些研究的结果通常以**描述统计**（descriptive statistics）的形式报告，描述被观察的被试所具有的特征。例如，研究者打算描述密歇根大学二年级学生的智力特征。收集到的数据（如IQ测试的结果）可以用平均分数和离差分数或图表的形式进行概括分析（心理学期刊论文中，同时采用统计数据和图表数据清晰呈现研究的结果）。

实验研究对一种或多种变量进行操纵，然后统计测量这些变量的结果。在实验和观察研究中，都可以使用另一种类型的统计分析。**推论统计**（inferential statistics）是一种搜集样本信息，并基于

该样本进行推论的数学方法。例如，有人研究一种草莓味早餐谷物食品在俄亥俄东北部的广告活动的效果，其目的是将这些结果推论到一个更大的总体（例如全国）。

统计。使用**统计检验**（statistical testing）对假设进行检验是当代实验心理学的重要部分。先前的章节中已经涉及不少这一方面的问题，在本书的第二部分大家还会看到很多统计的应用。确定某种特殊的数据类型适用何种统计检验方法是统计课程中的一个主题，具备了统计检验及其假设的基本知识之后，任何人都能够将正确的检验方法合理地应用于正确的条件。一旦决定了所要采用的统计方法，就可以非常容易地使用计算机和统计软件快速地进行大量的计算了。这些统计软件可以帮你计算出函数（F、t、χ^2 等）以及其他更复杂的统计过程。你甚至不需要知道统计中涉及的运算步骤，输入数据，结果就被输出了！不过，研究者必须要能够根据分析数据的类型，选取并运行适当的统计检验方法。

直条图和直方图。数据可以用图形的方式呈现，**直条图**（bar graph）和直方图就是这样的两种方法。这两种方法都使用了形象的图形，读者能够从中快速获得对实验结果的印象。

一般来说，被操纵的因素应放在**横轴**（abscissa）上，而结果则显示在**纵轴**（ordinate）上。

直方图（histogram）是根据频数表建构的。X 轴上显示各间距，而各间距中的频数则用该间距上直条的高度表示。直方图的数据是量化且连续的，直方图上的条形是相互连接的。而直条图则通常用于表示定性的分类数据，各直条互不相连。

案例分析

这里呈现的是戴蒙德（Diamond）和加里（Carey）的一个实验（参见图 9—2）中的直条图。他们对面孔的再认记忆感兴趣，并研究当面孔倒置时对再认的影响。作为补充，他们还评价了狗面孔的再认记忆。被试分为专家（饲养员、训练员和美国养犬俱乐部成员）和新手，研究者给他们呈现人和狗的面孔照片。如图所示，对狗面孔的记忆和人脸再认一样容易受到倒置的影响；另外，新手可能会由于缺乏足够的狗面孔的经验，因而不受面孔倒置的影响。换句话说，由于缺乏相关经验，他们可以把注意力放在狗面孔的其他特征上，因而即使在倒置时也能进行识别。

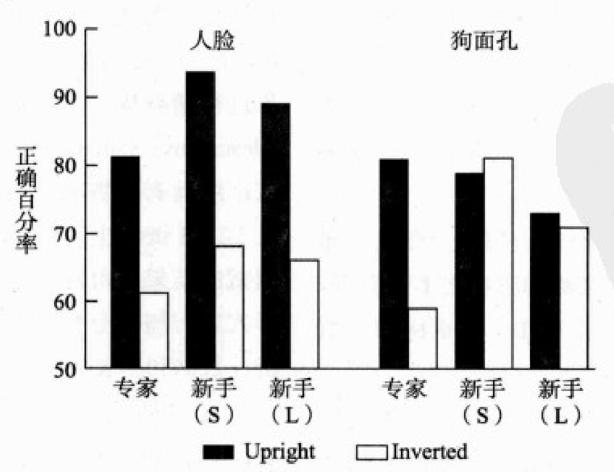


图 9—2　专家和新手在直立与倒置的人和狗面孔识别时的成绩（直条图）。新手（S）接受小样本的狗面孔刺激，而新手（L）和专家都接受相同的大样本的狗面孔刺激

图形和函数。图形可以描绘两个或多个变量间的关系。例如，当研究者得到一个因变量及其他几个变量（通常是自变量）不同水平的数值时，他就可以轻易绘出一条曲线。研究者可以从曲线图中看出两个变量的关系。例如，通过不同学习曲线的比较能够确定某种强化对条件学习的影响。

由图形还能得出一个描述两个变量关系的数学等式。这一等式也称做**函数**（function），是描述关系的便捷方法。一旦得出函数，就可以通过输入一个变量的数值，而获得另一个变量的数值。画曲线时，通常是将因变量置于纵轴，而将自变量置于横轴，从原点处开始随着一个变量数值的变化而变化。下列案例中的图（参见图 9—3）遵循了这一程序，图中的曲线清晰描绘了两个变量间的关系。

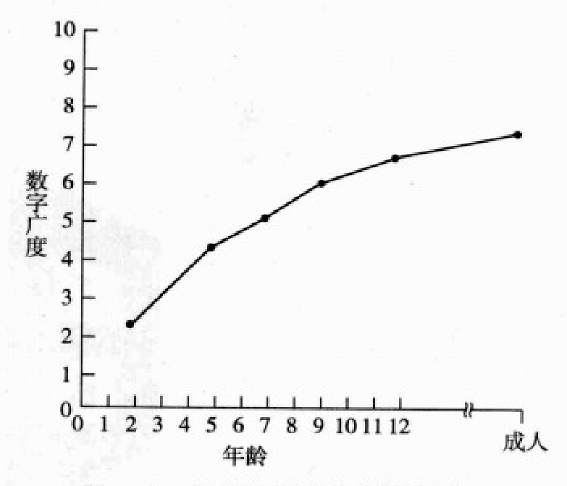
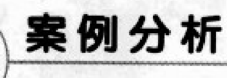


图 9—3　数字记忆随年龄变化的函数



案例分析

记忆研究通常使用的任务之一是对数字进行记忆——数字广度测试。数字广度测试要求被试复述先前呈现的一系列数字。如果研究者呈现的是“4-8-3-6-9”，被试的任务就是重复这一数字序列。

记忆研究中的一个有趣问题是：数字记忆广度是否是一个随着年龄而增长的函数，如果是，这种关系是怎样的？

在登普斯特（Dempster，1981）的研究中，不同年龄的儿童接受了一个数字广度测试。实验结果如图 9—3 所示，数字广度随着年龄而增长。这一现象的原因更加让人感兴趣，其中有一种解释是年长被试拥有更有意义的信息组块能力。年长的被试可能会将一个数字序列组块为两个数列（4-8 和 3-6-9），这会减少信息的负荷。所有人都会使用这种记忆的小窍门，有时甚至是无意识地使用这些窍门。

插图。除了使用统计方法外，研究者还可以使用各种视觉形式呈现研究的信息。有时，这些图形的呈现形式要优于文字。例如在以下案例分析中应用的照片和图片。

在这一案例分析中，一个图片和几张照片能够让读者很好地掌握研究使用的技术，并重现实验的程序。

研究者发现猴子能够辨别视觉刺激，例如面孔；因而他们认为灵长类的视觉记忆可能是以视觉编码的方式表征的。

案例分析

桑兹（Sands）、林肯（Lincoln）和莱特（Wright）在 1982 年研究了猕猴的视觉记忆。一般认为猴子不具备对物体的言语标签，这有助于回答视觉记忆加工的问题。研究者构建了如图 9—4 所示的装置，然后每次给被试同时呈现两幅照片，其中一张照片呈现在另一

张的上方（图中是香蕉和苹果）。研究者训练猴子向左或向右移动一个手柄，表示两个物体相同或不同。强化物是一个瓶子中的液体。实验使用的一些面孔如图 9—5 所示。

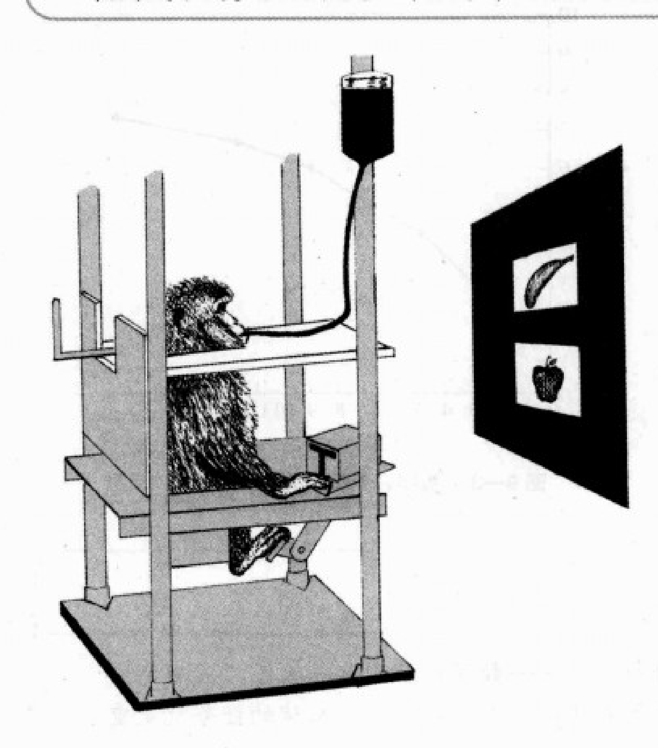


图 9—4　在桑兹等人的视觉记忆测验中，一只猕猴在操作一个杠杆装置



图 9—5　桑兹等人的视觉记忆研究中呈现给猕猴的部分面孔

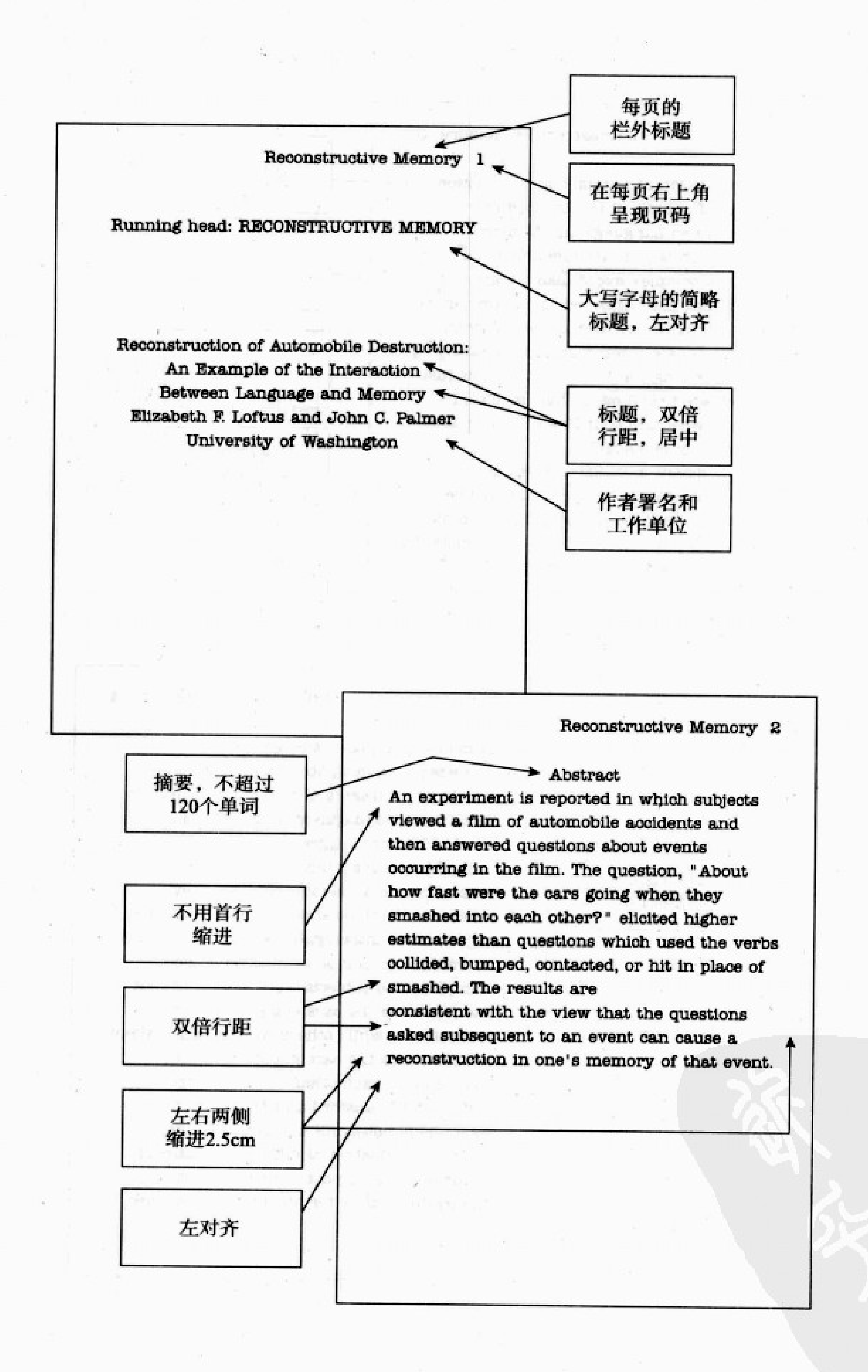
下一步我们将通过一个具体的样稿来了解如何在论文中报告实验的结果。

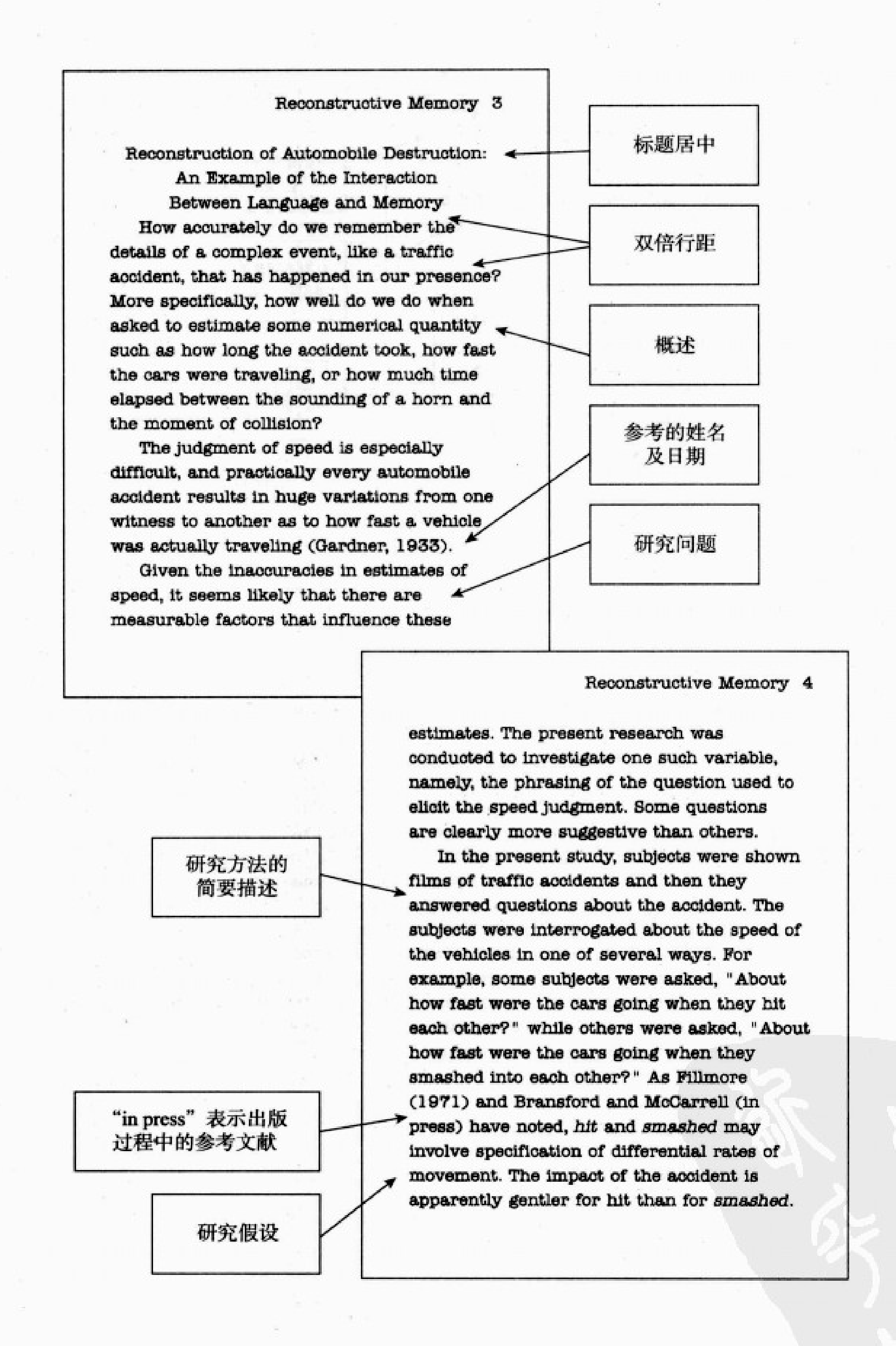
9.4　样稿

这部分包含一篇真正的研究论文，但相对于正式发表的版本略有修改。这篇论文（参见图 9—6）以准备提交给专业期刊的论稿形式呈现，提供了一个很好的实验设计模型，并说明了 APA 格式的合理应用。当你阅读这篇论稿时，请注意以下几个方面：（1）标题、作者、工作单位；（2）摘要；（3）文献综述、研究问题、假设；（4）研究方法（包括被试、仪器或材料、研究程序）；（5）研究结果（以数据和图表形式呈现）；（6）讨论和（7）参考文献。我们在论文中标识出了这些部分，但在提交论稿时是不存在这些标识的。

在洛夫特斯（Loftus）和帕尔默（Palmer）的这篇文章中，有几点需要特别注意。一个是摘要部分。这篇摘要有三个主要内容：告诉读者研究使用的方法；报告了基本的结论；用某种理论框架对结果进行解释。

另外要注意的一点是**文献综述**（literature review）部分。文章首先用一系列问题引发读者对研究问题的兴趣。这不是论文所必需的，但它在本实验中发挥了重要作用。接着作者呈现了其他研究者的几个发现，然后对研究问题进行描述。他们接着简要描述了研究的方法，并提出一个非正式的假设。在描述方法时，作者明确说明了研究的对象和程序；研究结果以文字、统计、图表方式呈现。讨论部分也是本论文中很有意义的一部分，作者简要回顾了研究发现，并提出了对数据的两种解释，接着，他们在已有研究和理论的基础上继续对结果进行讨论和解释；但他们的讨论并没有超出本实验结果的范围。最后一部分是参考文献，学生在列举自己写作中的参考文献时，可以拿这部分作为范本。


Reconstructive Memory 1
Running head: RECONSTRUCTIVE MEMORY
Reconstruction of Automobile Destruction:
An Example of the Interaction
Between Language and Memory
Elizabeth F. Loftus and John C. Palmer
University of Washington
每页的
栏外标题
在每页右上角
呈现页码
大写字母的简略
标题，左对齐
标题，双倍
行距，居中
作者署名和
工作单位
Reconstructive Memory 2
Abstract
An experiment is reported in which subjects viewed a film of automobile accidents and then answered questions about events occurring in the film. The question, "About how fast were the cars going when they smashed into each other?" elicited higher estimates than questions which used the verbs collided, bumped, contacted, or hit in place of smashed. The results are consistent with the view that the questions asked subsequent to an event can cause a reconstruction in one's memory of that event.
摘要，不超过
120个单词
不用首行
缩进
双倍行距
左右两侧
缩进2.5cm
左对齐



Reconstructive Memory 3

Reconstruction of Automobile Destruction:
An Example of the Interaction
Between Language and Memory

How accurately do we remember the details of a complex event, like a traffic accident, that has happened in our presence? More specifically, how well do we do when asked to estimate some numerical quantity such as how long the accident took, how fast the cars were traveling, or how much time elapsed between the sounding of a horn and the moment of collision?

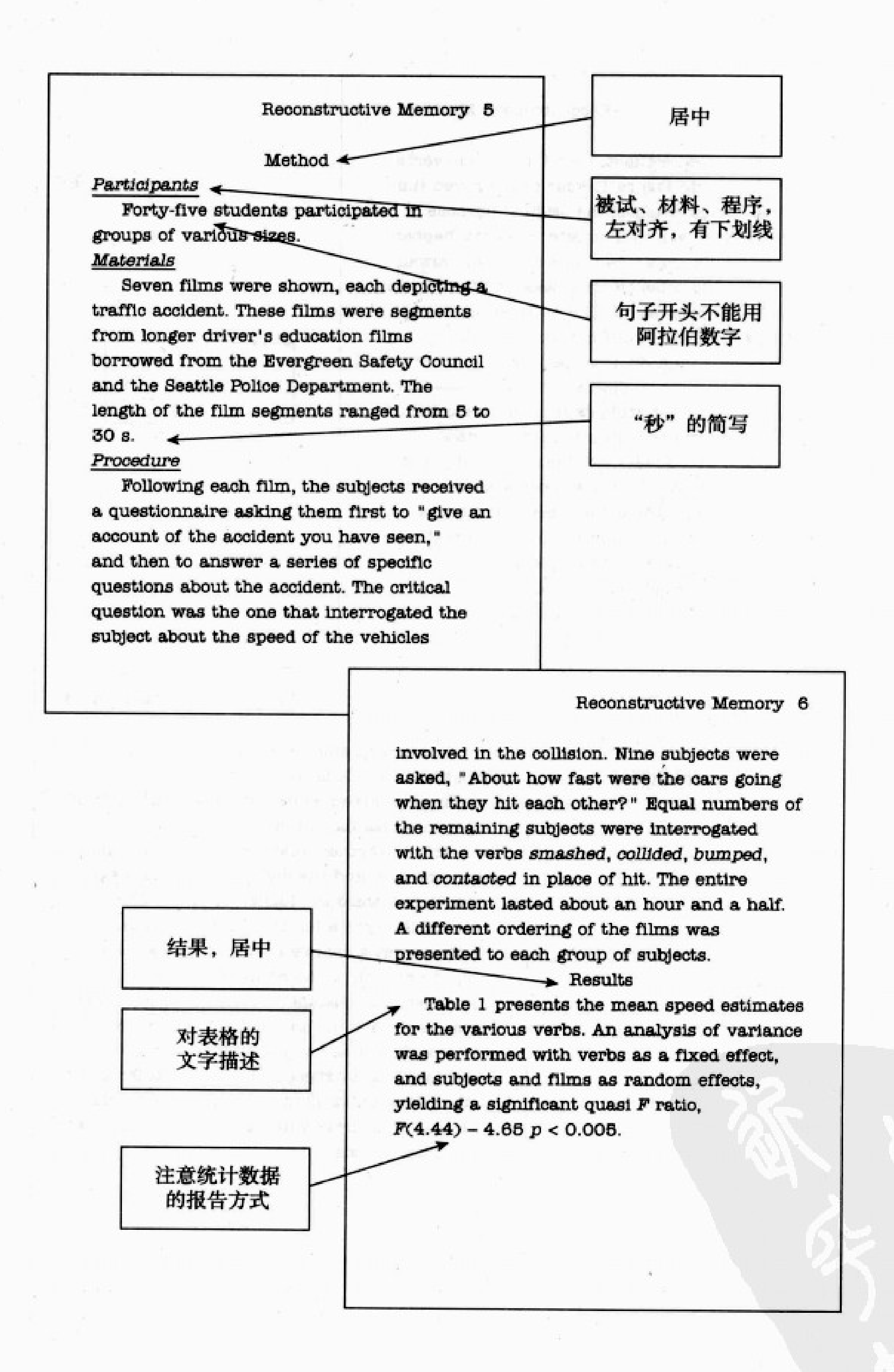
The judgment of speed is especially difficult, and practically every automobile accident results in huge variations from one witness to another as to how fast a vehicle was actually traveling (Gardner, 1933).

Given the inaccuracies in estimates of speed, it seems likely that there are measurable factors that influence these

Reconstructive Memory 4

estimates. The present research was conducted to investigate one such variable, namely, the phrasing of the question used to elicit the speed judgment. Some questions are clearly more suggestive than others.

In the present study, subjects were shown films of traffic accidents and then they answered questions about the accident. The subjects were interrogated about the speed of the vehicles in one of several ways. For example, some subjects were asked, "About how fast were the cars going when they hit each other?" while others were asked, "About how fast were the cars going when they smashed into each other?" As Fillmore (1971) and Bransford and McCarrell (in press) have noted, *hit* and *smashed* may involve specification of differential rates of movement. The impact of the accident is apparently gentler for hit than for *smashed*.



Reconstructive Memory 5

Method

Participants

Forty-five students participated in groups of various sizes.

Materials

Seven films were shown, each depicting a traffic accident. These films were segments from longer driver's education films borrowed from the Evergreen Safety Council and the Seattle Police Department. The length of the film segments ranged from 5 to 30 s.

Procedure

Following each film, the subjects received a questionnaire asking them first to "give an account of the accident you have seen," and then to answer a series of specific questions about the accident. The critical question was the one that interrogated the subject about the speed of the vehicles

Reconstructive Memory 6

involved in the collision. Nine subjects were asked, "About how fast were the cars going when they hit each other?" Equal numbers of the remaining subjects were interrogated with the verbs *smashed*, *collided*, *bumped*, and *contacted* in place of hit. The entire experiment lasted about an hour and a half. A different ordering of the films was presented to each group of subjects.

Results

Table 1 presents the mean speed estimates for the various verbs. An analysis of variance was performed with verbs as a fixed effect, and subjects and films as random effects, yielding a significant quasi F ratio, $F(4.44) - 4.65$ $p < 0.005$.

Reconstructive Memory 7

The speed estimates for the various verbs are shown in Figure 1. Four of the seven films were staged crashes; the original purpose of these films was to illustrate what can happen to human beings when cars collide at various speeds. One collision took place at 20 mph. one at 30, and two at 40. The mean estimates of speed for these four films were: 37.7, 36.2, 39.7, and 36.1, respectively.

Discussion

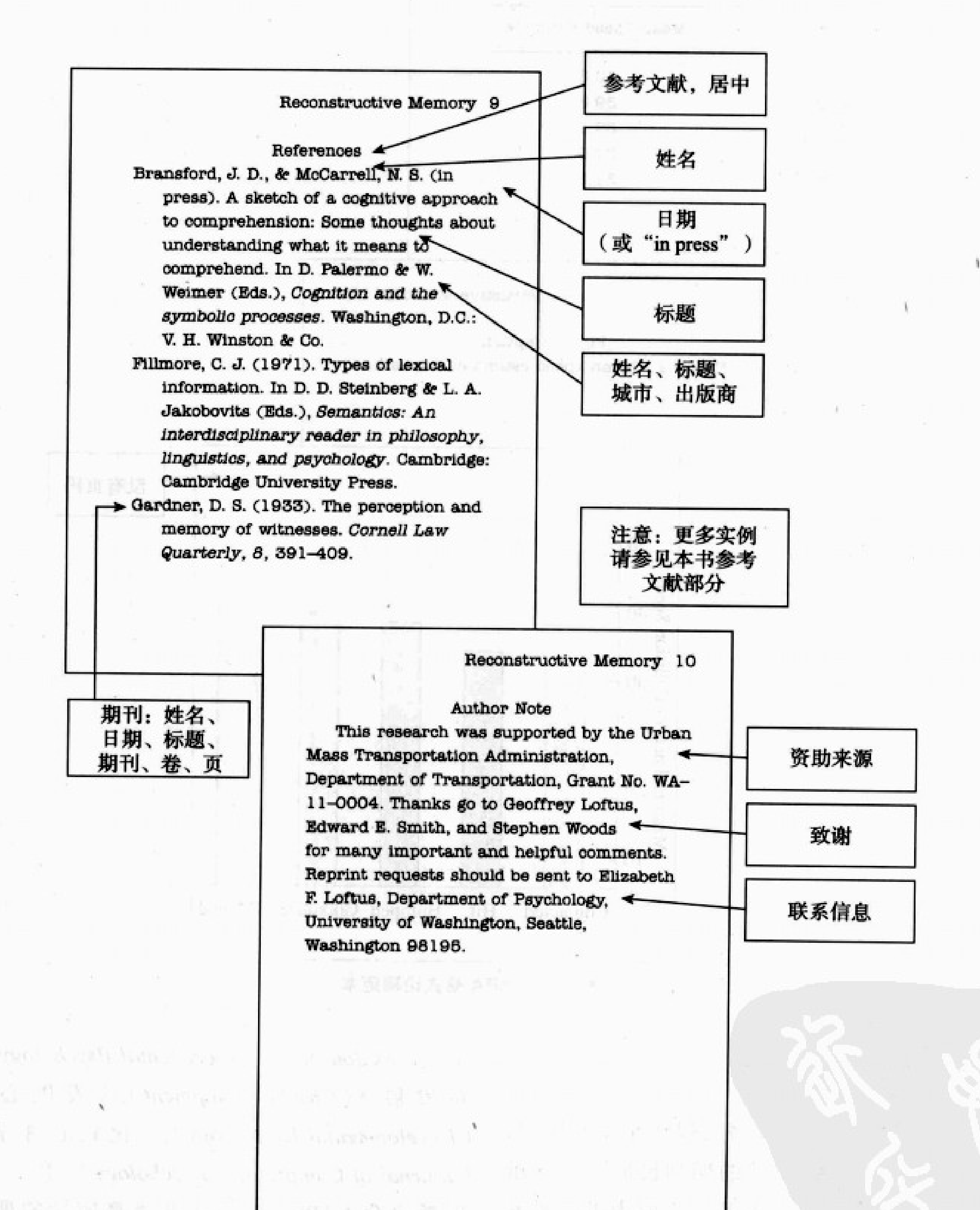
居中

The results of this experiment indicate that the form of a question (in this case changes in a single word) can markedly and systematically affect a witness's answer to that question. The actual speed of vehicles controlled little variance in subject reporting, while the phrasing of the question controlled considerable variance.

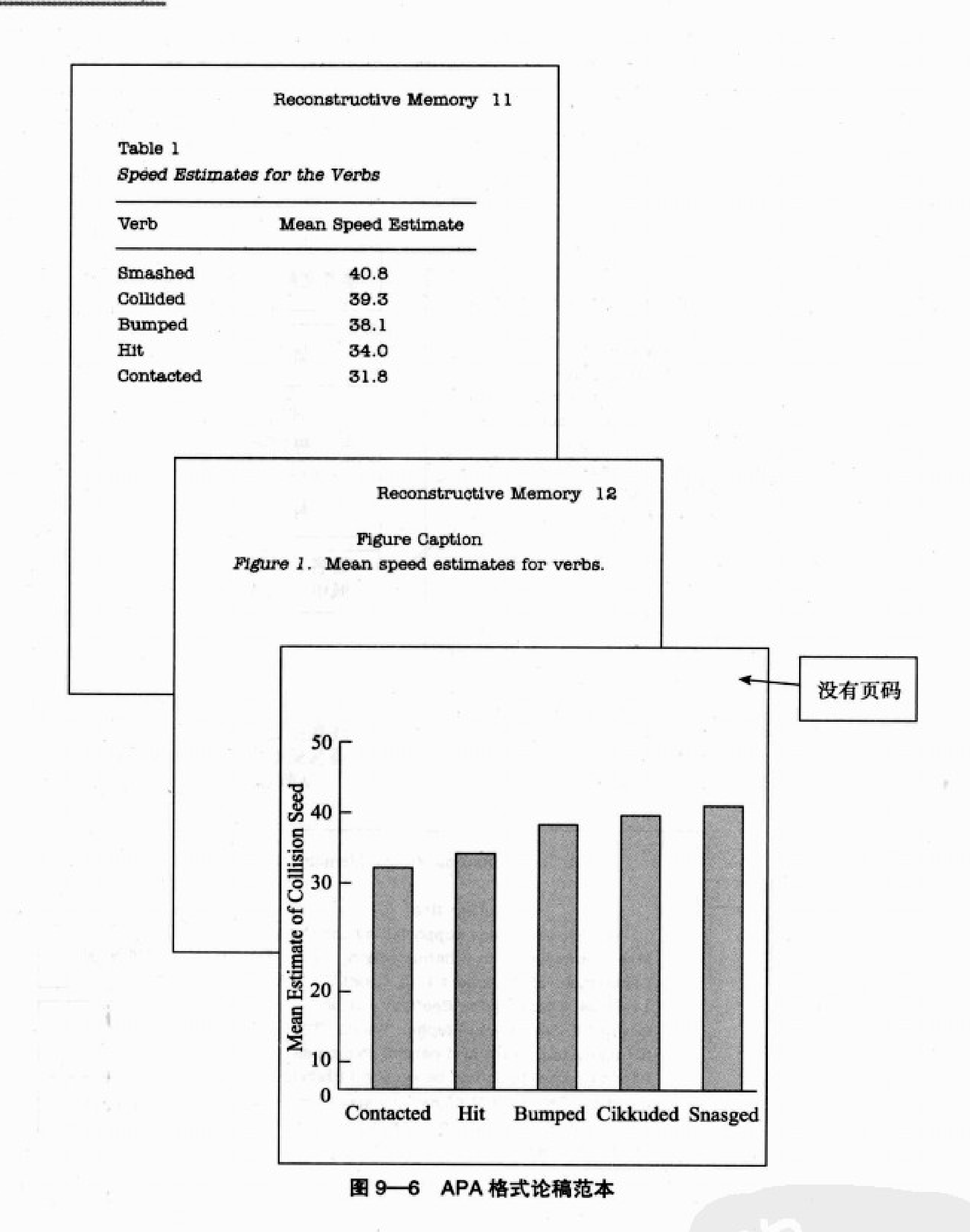
Reconstructive Memory 8

数据的解释

Two interpretations of this finding are possible. First, it is possible that the differential speed estimates result merely from response-bias factors. A subject is uncertain whether to say 30 mph or 40 mph, for example, and the verb smashed biases his response toward the higher estimate. A second interpretation is that the question form causes a change in the subject's memory representation of the accident. The verb smashed may change a subject's memory such that he "sees" the accident as being more severe than it actually was. If this is the case, we might expect subjects to "remember" other details that did not actually occur but are commensurate with an accident occurring at higher speeds.



Reconstructive Memory 9
References
Bransford, J. D., & McCarrell, N. S. (in press). A sketch of a cognitive approach to comprehension: Some thoughts about understanding what it means to comprehend. In D. Palermo & W. Weimer (Eds.), Cognition and the symbolic processes. Washington, D.C.: V. H. Winston & Co.
Fillmore, C. J. (1971). Types of lexical information. In D. D. Steinberg & L. A. Jakobovits (Eds.), Semantics: An interdisciplinary reader in philosophy, linguistics, and psychology. Cambridge: Cambridge University Press.
Gardner, D. S. (1933). The perception and memory of witnesses. Cornell Law Quarterly, 8, 391–409.
参考文献，居中
姓名
日期
（或“in press”）
标题
姓名、标题、
城市、出版商
注意：更多实例
请参见本书参考
文献部分
期刊：姓名、
日期、标题、
期刊、卷、页
Reconstructive Memory 10
Author Note
This research was supported by the Urban Mass Transportation Administration, Department of Transportation, Grant No. WA-11-0004. Thanks go to Geoffrey Loftus, Edward E. Smith, and Stephen Woods for many important and helpful comments. Reprint requests should be sent to Elizabeth F. Loftus, Department of Psychology, University of Washington, Seattle, Washington 98195.
资助来源
致谢
联系信息



Reconstructive Memory 11

Table 1

Speed Estimates for the Verbs

Verb	Mean Speed Estimate
Smashed	40.8
Collided	39.3
Bumped	38.1
Hit	34.0
Contacted	31.8

Reconstructive Memory 12

Figure Caption

Figure 1. Mean speed estimates for verbs.

图 9—6　APA 格式论稿范本

9.5　投稿

论稿完成后，作者需要将之投送至专业期刊，以便审稿和出版。选择合适的期刊投稿是件困难的事情，对新手来说尤其如此。一般来说，研究者们更愿意把自己的作品发表在知名的、权威的期刊上。当然，刊物的“好”与“不好”在于你的理解。APA 有几种叫好又卖座的实证研究的刊物，例如《人格与社会心理学杂志》(*Journal of Personality and Social Psychology*)、《实验心理学杂志》(*Journal of Experimental Psychology*)、《儿童发展》(*Child Development*)、《发展心理学》(*Developmental Psychology*)、《比较心理学杂志》(*Journal of Comparative Psychology*)等。美国心理学协会（APS）出版了几类高质量的期刊，例如《心理科学最新动态》(*Current Directions in Psychological Science*)。许多团体也出版了很多重要的期刊，例如基础心理协会（*Psychonomic Society*）出版了《记忆和认知》(*Memory & Cognition*)、《知觉与心理物理学》(*Perception & Psychophysics*)

和《基础心理协会会刊》(*Bulletin of the Psychonomic Society*)，英国心理学协会（BPS）出版了《英国心理学杂志》(*British Journal of Psychology*) 和《实验心理学季刊》(*Quarterly Journal of Experimental Psychology*)。另外还有许多期刊可供选择，如《美国心理学杂志》(*American Journal of Psychology*)、《认知心理学》(*Cognitive Psychology*)、《神经心理学》(*Neuropsychology*)、《脑与认知》(*Brain and Cognition*)、《大脑皮层》(*Cortex*)、《行为的实验分析》(*Journal of Experimental Analysis of Behavior*)、《记忆和语言杂志》(*Journal of Memory & Language*)等。

以上仅列举了心理学期刊的一部分，还有许多高水平的刊物出版特定主题的实验心理学研究。研究者应当选取最适宜的专业刊物投稿。你可以通过文献阅读和综述，找到那些同类问题的研究者们所投稿的刊物。通常刊物都会有一个或几个一贯性的主题。

在期刊投稿之后，即使许多编辑希望能够在三个月内定稿，但实际上整个审稿的过程可能会花上一年。图 9—7 简要描述了 APA 的编辑过程。

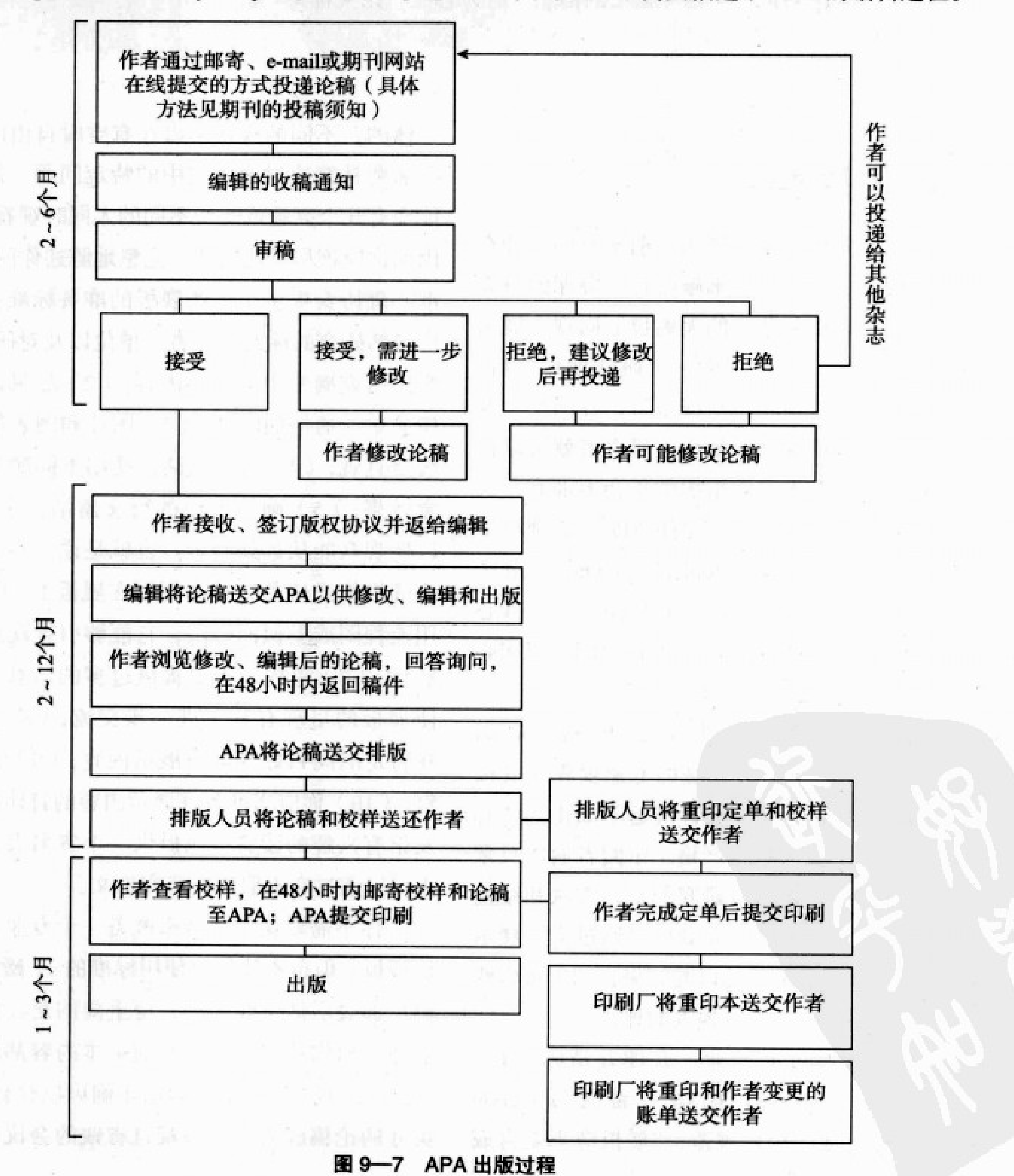


图 9—7　APA 出版过程

专栏

撰写摘要

摘要是对你的研究项目的简要总结，通常英文字数在 300 ~ 1 000 个单词之间（参照会议提出的长度标准）。摘要应当包括：清晰明确的标题、对研究主题的简要介绍、对研究目标的陈述、对研究方法的呈现、对研究结果的总结以及研究的结论。有些会议还允许提交结果尚不明了的研究摘要（因为你的研究可能还在进行中，但能够在会议开始前完成），但其他一些会议则不允许这种情况。因此，你需要仔细阅读关于论文提交的信息（有时称做“论文征集”）。

9.6　专业会议的展板

专业会议期间的展板是展示当前研究的一种有效方式，这种方式正变得越来越流行。资深的研究者和学生都可以直接在独立的大展板上向观众展示自己的研究。通常，当你完成一个研究项目之后，需要进行必要的统计检验，写成报告，并向某个专业组织提交摘要。摘要被接受后，研究者就可以在会议上进行展示了。大多数组织的年会上都有会议展板环节，并鼓励学生参与，比如 APA、心理科学协会（Association for Psychological Science）、地区性组织（东部心理协会、中西部心理协会、西部心理协会、落基山心理协会等）、专业性组织（美国心理学与法学协会、行为分析者协会）等。

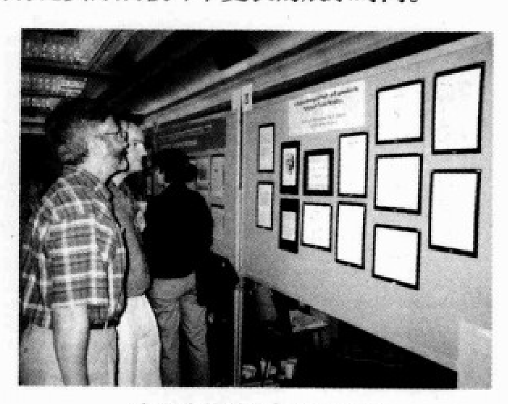
一篇论文在被接受为**会议展板**（poster presentation）之前，该研究领域的专家要先对其摘要进行审阅。会议展板的接受标准远不如正式的心理学期刊接受全文研究报告严格。审阅者通常只拿到研究的摘要，因而无法对研究的具体方法和结论进行评价。正因为如此，研究者可以通过会议展示的方式（1）只报告一部分的研究结果，（2）报告暂时性的结论，或是（3）获得他人的评论。

会议展板应尽可能专业、有趣并清晰明确。人格与社会心理协会的托德·希瑟顿（Todd Heatherton）提出他的展板规范：“展板应当是自成一体的，不同的观众可以在有空时自由阅读。作者只需要补充或讨论研究中的特定问题。请注意，可能会有几个兴趣或经历不同的人同时观看你的展板，因而你应该尽可能简洁、完整地陈述你的观点。”西部心理协会提供了一些展板的准备标准：（1）展板应包括研究的标题、作者、单位以及对研究的描述，突出研究摘要中的主要内容；（2）尽量详细，并使用简单、通俗的陈述；（3）图片和图表能使你的展板更直观；（4）可能的话，使用不同颜色来增加视觉效果；（5）确信展板的行文整洁，字号足够大，以便观众能从远处阅读。也就是说，不要仅仅是把你的提纲或传单的打印页钉在展板上；（6）考虑使用流程图或类似的方法，它能够引导观众阅读整个展板；（7）首先，不要提供过多的信息；另外，应使展板的呈现有利于进一步交流；（8）做好准备，在特定时间钉好和取下展示内容；（9）记得带上图钉；（10）你应当准备好展示内容的打印副本，以便发给有兴趣的读者；或提供一个签名表，感兴趣的人可以通过它索取你的研究论文。

你不需要花很多钱来准备一个专业、有效的会议展板。但也不能仅仅使用标准的 12 磅字的打印论稿！那表示你不够用心，更重要的是，它不能够有效地突出你的研究。所有的文本内容都应当足够大（至少 26 磅字），使用彩色印刷可以让你的 8.5×11 英寸的论稿成为一个美观且省钱的会议展板。在保

证专业性的前提下，你可以在展板上添加一些便签、横幅、学校贴标等。另外，你也可以用小号字在单页上建构你的展板，然后在复印店或打印店放大即可。酒店商务中心的费用会比较高，考虑到你的神经和荷包，你最好在抵达会议之前准备好展板和准备发放的提纲。总之，你的展板形式要能够引起大家的兴趣，易于理解，并能够使人留下深刻印象。

你通常会分派到一个 4 英尺高、8 英尺宽的展板。不过，你最好详细查阅会前材料，了解该会议的具体细则和任何特殊的要求。纵向的网格方式是比较有效的展板排列模式，这可以避免观众要不断地走动来阅读你的展板。你的展板展示的时间通常为 2 个小时左右；记住，这是你和大家会面，并展示研究的机会，所以你最好能够在场。有些会议有额外的观看时间，你会有比实际展板环节更长的展示时间。



专业会议的展板展示时间

会议展板的目的是为了给新研究提供展示的场所。观众们围绕在展板周围，阅读展板的内容，向研究者咨询问题或进行评论。这通常是热烈的思想交流的场所；你可以和许多对你的研究感兴趣的人们进行面对面的交流，因而你应该准备充分、保持清醒的思维和友好的态度。注意，如果正在和他人进行与展板内容无关的友好交谈，那么许多对你的研究感兴趣的人就会却步了。要时刻留意展板周围的观众，引领他们浏览你的展板，或解答他们的疑问。另外要注意的一点是：专业着装。如果你对此有任何疑问，请在与会前咨询你的挚友、导师或是教授。要问我们的建议是什么？那就是祝你愉快！毕竟，你为该研究付出了很大的努力，而现在正是你放松心情，享受这一成果的时刻。

9.7　结束语

至此，你已经完成了本书方法部分的学习。下面，你将会通过几个真实研究报告的阅读，进一步验证并扩展实验设计的知识。



[illegible]

[illegible]

第二部分

实验分析

本书第二部分通过精心选择的真实文献的分析，说明心理学的各种实验技术和主题。

心理学家在完成一项实验后，下一个任务就是报告其实验的结果。也就是说，心理学家要准备一篇用于发表的研究论文。因此，理解性阅读及撰写这类文章的能力是对研究型心理学家的最重要的要求之一。

由于期刊、杂志的版面有限，科学家需要花费一定的时间处理论文，清晰准确、简明扼要是呈现研究结果的关键部分。同时，由于阅读论文的是其他科学家，尤其是研究同一问题或领域的科学家，因而作者还须让读者了解该领域的研究历史和研究方法。杂志论文的撰写并非一蹴而就，不同研究方向的学生或心理学者都将要面对这一困难。

第二部分中的某些论文经过了一定的编辑处理；另一些则按其全文呈现。在这些论文中，你可以发现如何去应用所学的实验设计的知识。大多数文章还伴有本书作者的分析。通过对这类文章的不断接触和细致分析，学生们会逐渐习惯阅读研究报告，并对心理学实验设计原则有更好的理解。有些论文没有进行文献分析，你可以尝试这一学习过程。

文献分析是依照原文中的各部分逐节进行的。我们把分析部分放在原论文相应部分的下面。我们选择的许多实验都能够说明某种研究主题，这些研究主题一般在文献的开头，独立于我们所作的分析。下列的各项主题都至少有一篇研究文献作为例子。虽然对研究文献的阅读并无唯一的标准，但我们接下来给出的建议也许能让你更好地理解研究报告。